

# Robust block bootstrap panel predictability tests

Citation for published version (APA):

Westerlund, J., & Smeeke, S. (2013). *Robust block bootstrap panel predictability tests*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 060 <https://doi.org/10.26481/umagsb.2013060>

**Document status and date:**

Published: 01/01/2013

**DOI:**

[10.26481/umagsb.2013060](https://doi.org/10.26481/umagsb.2013060)

**Document Version:**

Publisher's PDF, also known as Version of record

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Joakim Westerlund,  
Stephan Smeekes

**Robust Block Bootstrap Panel  
Predictability Tests**

RM/13/060

**GSBE**

Maastricht University School of Business and Economics  
Graduate School of Business and Economics

P.O Box 616  
NL- 6200 MD Maastricht  
The Netherlands

# ROBUST BLOCK BOOTSTRAP PANEL PREDICTABILITY TESTS\*

Joakim Westerlund<sup>†</sup>

Deakin University  
Australia

Stephan Smeekees

Maastricht University  
The Netherlands

August 2, 2013

## Abstract

Most panel data studies of the predictability of returns presume that the cross-sectional units are independent, an assumption that is not realistic. As a response to this, the current paper develops block bootstrap-based panel predictability tests that are valid under very general conditions. Some of the allowable features include heterogeneous predictive slopes, persistent predictors, and complex error dynamics, including cross-unit endogeneity.

**JEL Classification:** C15; C22; C23; G1; G12.

**Keywords:** Panel data; Predictive regression; Stock return predictability; Sequential testing; Block bootstrap.

## 1 Introduction

Consider the panel data variable  $y_{i,t}$ , observable for  $t = 1, \dots, T$  time series and  $i = 1, \dots, N$  cross-sectional units. Recent years have witnessed an immense proliferation of research asking whether  $y_{i,t}$  can be predicted using the one-period lagged value of some other variable,  $x_{i,t}$  say. Examples of such situations are abound. The most common ones are found in finance. For example, if  $y_{i,t}$  is stock returns, or the equity risk premium, then  $x_{i,t}$  might be

---

\*The first author would like to thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2005-0117:1. The second author would like to thank the Netherlands Organization for Scientific Research (NWO) for financial support.

<sup>†</sup>Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia. Telephone: +61 3 924 46973. Fax: +61 3 924 46283. E-mail address: j.westerlund@deakin.edu.au.

dividend yield, nominal interest rates, default or term spreads on bonds, inflation, valuation ratios, the consumption-wealth ratio, stock market volatility, labor income, aggregate output, output gap, or oil prices, just to mention a few (see Neely et al., 2012; Rapach and Zhou, 2012, and the references provided therein).

The conventional way in which earlier studies have been trying to test the predictability hypothesis is to first run a time series regression of  $y_{i,t}$  onto a constant and  $x_{i,t-1}$ , and then to test whether the (predictive) slope on  $x_{i,t-1}$  is zero by using a conventional  $t$ -test.<sup>1</sup> This test is then repeated for each unit in the sample, each time using only the sample information for that particular unit (see, for example, Ang and Bekaert, 2007; Driesprong et al., 2008; Polk et al., 2006; Rapach et al., 2012).

Hjalmarsson (2010) questions the unit-by-unit approach and suggests combining the sample information obtained from the time series dimension with that obtained from the cross-sectional (see also Hjalmarsson, 2008; Kauppi, 2001). Specifically, a pooled  $t$ -test is proposed, which not only increases the power by taking the total number of observations and their variation into account, but also increases the precision of the predictability test by effectively reducing the noise coming from the individual time series regressions. Moreover, where the unit-by-unit approach is likely to lead to an over-rejection of the no predictability null, the panel approach accounts for the multiplicity of the testing problem and is therefore correctly sized.

But while the approach of Hjalmarsson (2010) has many advantages, it also has its fair share of drawbacks. The first problem is the formulation of the hypothesis tested. In particular, while the null hypothesis can certainly be formulated as that there is no predictability, the alternative hypothesis that there are at least some units for which predictability holds is too broad for any interesting economic conclusions; it could be that there is predictability for all units, but it could also be that there is only a small fraction of units for which predictability holds. Another problem is the way in which the test is made robust to cross-section dependence. Specifically, a common factor structure is assumed, the effect of which, following the common practice in the panel unit root literature (see Bai and Ng, 2010, page 1097, for a discussion), is removed prior to implementation of the test for predictability. Hence, with this approach one is essentially testing for predictability in the remaining idiosyncratic compo-

---

<sup>1</sup>Some of the many time series tests that have been proposed include Stambaugh (1999), Lanne (2002), Lewellen (2004), and Campbell and Yogo (2006), to mention a few.

nent, thereby ignoring a potentially important source of predictive information, namely, the common one. Then there is also the fact that the assumed common factor structure need not be correct, and, even if it is, for the (cross-section dependence robust) test to be valid the predictive slopes are still restricted to be homogenous (see Hjalmarsson, 2010, Theorem 6 and Corollary 1), an assumption that is certainly mistaken in practice. Finally, there is also the requirement that  $N$  should go to infinity with  $T$ , which may be motivated in applications involving highly disaggregated panels, but not in the typical large- $T$ , small- $N$  (financial) cross-country study (see, for example, Rapach et al., 2012).

In this paper, we develop several procedures to ascertain the predictability of a panel. The point of departure is a very general data generating process (DGP) that allows, for example, heterogeneous predictive slopes, persistent predictors, and complex error dynamics, including cross-unit endogeneity. In fact, except for some mild regulatory conditions, there are virtually no restrictions on the forms of serial and cross-sectional dependence that can be permitted. Given this generality, corrections aimed at achieving asymptotically pivotal statistics are not really an option. In this paper we therefore consider the block bootstrap as a means to obtain tests that are asymptotically valid. In doing so, we further the recent work of Palm et al. (2011) for univariate unit root panels to a bivariate model. Two block bootstrap-based test procedures are considered; one is appropriate when testing the above mentioned hypothesis of full panel unpredictability versus at least some predictability, while the other can be used to sequentially determine the units for which predictability holds. Both procedures do not require letting  $N$  go to infinity, and in fact perform well even when  $N$  is relatively small.

The rest of the paper is organized as follows. Section 2 introduces the model. The test statistics and their bootstrap implementations are presented in Section 3. The asymptotic properties of the methods are analyzed in Section 4. In Section 5 a Monte Carlo simulation study is presented. Section 6 concludes. Proofs are provided in Appendix.

## 2 The model

Consider the  $N$ -dimensional panel data variables  $y_t = (y_{1,t}, \dots, y_{N,t})$  and  $x_t = (x_{1,t}, \dots, x_{N,t})'$ . The DGP of these variables is given by

$$y_t = \alpha + \beta x_{t-1} + v_t, \quad (1)$$

$$x_t = \delta(1 - \rho) + \rho x_{t-1} + w_t, \quad (2)$$

where  $x_0 = 0$ ,  $\alpha = (\alpha_1, \dots, \alpha_N)'$ ,  $\beta = \text{diag}(\beta_1, \dots, \beta_N)$ ,  $\delta = (\delta_1, \dots, \delta_N)'$  and  $\rho = \text{diag}(\rho_1, \dots, \rho_N)$ . This is a panel extension of the prototypical predictive regression model that has been widely used in the time series literature, in which  $x_t$  is a variable believed to be able to predict  $y_t$ . According to this literature, it is reasonable to assume that  $v_t$  is correlated with  $w_t$ . For example, if  $x_t$  is returns and  $x_t$  is the dividend-price ratio, then an increase in the stock price will lower dividends and raise returns. We therefore assume that

$$u_t = \Psi(L)\varepsilon_t, \quad (3)$$

where  $u_t = (v_t', w_t')'$  and  $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$  with  $\Psi_0 = I_{2N}$ . The rest of the assumptions are as follows.

### Assumption 1.

- (a)  $\sum_{j=0}^{\infty} j \|\Psi_j\| < \infty$  and all the rows of  $\Psi(1)$  are nonzero;
- (b)  $\varepsilon_t$  is independently and identically distributed (iid) with  $\mathbb{E} \varepsilon_t = 0$ ,  $\mathbb{E} \varepsilon_t \varepsilon_t' = \Sigma_{\varepsilon\varepsilon}$  and  $\mathbb{E} \|\varepsilon_t\|^\kappa < \infty$  for some  $\kappa \geq 4$ .

### Assumption 2.

$$\begin{aligned} \rho &= 1 + \frac{cm}{T}, \\ \beta &= \frac{bm}{T}, \end{aligned}$$

where  $m > 0$  is a scalar such that  $m \rightarrow \infty$  as  $T \rightarrow \infty$ ,  $c = \text{diag}(c_1, \dots, c_N) < 0$  and  $b = \text{diag}(b_1, \dots, b_N)$ .

Under Assumption 1, the long-run covariance matrix of  $u_t$  is given by

$$\Omega = \lim_{T \rightarrow \infty} T^{-1} \mathbb{E} \left( \sum_{t=1}^T u_t \right) \left( \sum_{t=1}^T u_t \right)' = \begin{bmatrix} \Omega_{vv} & \Omega_{vw} \\ \Omega_{wv} & \Omega_{ww} \end{bmatrix} = \Psi(1) \Sigma_{\varepsilon\varepsilon} \Psi(1)' = \Sigma + \Lambda + \Lambda',$$

where

$$\begin{aligned}\Sigma &= \mathbb{E} u_t u_t' = \sum_{j=0}^{\infty} \Psi_j \Sigma_{\varepsilon\varepsilon} \Psi_j', \\ \Lambda &= \mathbb{E} \sum_{k=1}^{\infty} u_t u_{t+k}' = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \Psi_j \Sigma_{\varepsilon\varepsilon} \Psi_{j+k}'\end{aligned}$$

are the contemporaneous and one-sided long-run covariance matrices of  $u_t$ , respectively, which are partitioned conformably with  $\Omega$ . In what follows,  $\Omega_{vv}$  and  $\Lambda_{vv} = \mathbb{E} \sum_{k=1}^{\infty} w_t v_{t+k}'$  are going to be particularly important, and we are therefore going to use  $\omega_{v,i}^2 = [\Omega_{vv}]_{ii}$  and  $\lambda_{vv,i} = [\Lambda_{vv}]_{ii}$ , respectively, to denote their diagonal elements.

**Remark 1.** The assumption placed on  $\rho$  ensures that  $x_t$  is “weakly integrated” (Park, 2003), although not “local-to-unity”, as when  $m$  is fixed.<sup>2</sup> The reason for this assumption is that the validity of the bootstrap depends critically on whether or not the model can be estimated while at the same time permitting continuity of the asymptotics. For the DGP considered here, the asymptotics are continuous in  $m/T$ -neighborhoods (with  $m, T \rightarrow \infty$  and  $m/T \rightarrow 0$ ) but not in  $1/T$ -neighborhoods, suggesting that while consistent in the weakly integrated case, in the local-to-unit case the bootstrap is inconsistent (see, for example, Stock, 1997, page 55; Park, 2006, page 640). Exact unit roots ( $c = 0$ ) are also excluded. While these assumptions could be considered a limitation, we do not believe this to be the case. There are at least two reasons for this. First, local and weak unit roots are just asymptotic concepts designed to capture relevant features found in practice when variables are highly persistent yet not unit root non-stationary. They therefore explain the same phenomena, and as such choosing one over the other does not necessarily imply a restriction (at least not in finite samples). Indeed, as Park (2006) argues, the weak unit root framework provides a good description of observed behavior even when the true DGP is local-to-unity. Second, even if many predictors can be highly persistent, there is little evidence to suggest that they have an exact unit root (see, for example, Lewellen, 2004; Westerlund and Narayan, 2012).

**Remark 2.** Wolf (2000) considers subsampling for the purpose of predictability testing, but, like us, has to exclude exact unit roots in order to show the validity of his subsampling approach. An importance difference when compared to the current weak unit root assumption

---

<sup>2</sup>The weak integration assumption is similar in spirit to the “moderate integration” assumption of Giraitis and Phillips (2006), Phillips and Magdalinos (2007), Phillips and Magdalinos (2009), and Phillips et al. (2010). However, in this paper, in order to facilitate the use of the bootstrap, we follow Park (2006) and assume that  $x_t$  is weakly integrated.

is that Wolf (2000) assumes that  $|\rho_i| < 1$  is fixed, and it is unknown whether his approach is valid in the present more general setting.

**Remark 3.** If one believes that  $x_t$  contains unit roots, then this restriction can be imposed in the bootstrap scheme. Such a strategy is commonly pursued in cointegration analysis, and it is not hard to show that the resulting bootstrap is valid. Unfortunately, this is only possible if the unit root restriction is indeed true. Allowing for both near and exact unit roots is considerably more complicated. Extensions of the current approach to cover also such cases are currently under investigation by the authors.

**Remark 4.** Since the properties of the bootstrap in stationary autoregressions are well known (see, for example, Bose, 1988; Künsch, 1989), our bootstrap approach can be easily adapted to cover the case when  $|\rho_i| < 1$  is fixed. The problem is that if  $|\rho_i| < 1$  is fixed, the presence of endogeneity will render OLS inconsistent, and this is true regardless of whether the bootstrap is used or not. However, given the persistence of many predictor candidates, this should not be too much of an issue in the current context.

**Remark 5.** Unlike  $c$ ,  $b$  is not restricted to be different from zero, which means that the predictability can be both “weak” ( $c \neq 0$ ) and absent ( $c = 0$ ). However, in analogy to the discussion in Remark 1 above, the predictability cannot be “local”, in the sense that  $m$  cannot be fixed. Thus, while the predictability can be made arbitrarily weak, to achieve non-trivial power the rate at which  $\beta \rightarrow 0$  has to be slower than when  $\beta$  is local-to-zero.

**Remark 6.** Except possibly for the relatively slow rate of shrinking (see Remarks 1 and 5 above), Assumption 2 is very general when it comes to types of predictability and persistency behaviors that can be permitted. Note in particular how the elements of  $c$  ( $b$ ) may differ, which means that the extent of the predictability (persistency) may vary across the cross-section. In fact, we could even allow  $m$  to vary across  $i$ , suggesting that the rate at which  $\rho \rightarrow 1$  ( $\beta \rightarrow 0$ ) need not be the same.

**Remark 7.** The types of cross-sectional dependencies that can be accommodated within the current DGP are more general than those that have been considered earlier in the literature (see Hjalmarsson, 2010; Kauppi, 2001), and include Granger causality, common factors and even “weak cointegration” between units. One way to think about the concept of weak



cointegration is, in analogy with common factor induced cointegration in unit root panels, that the persistency of  $x_{i,t}$  is driven by a weak unit root common factor. By allowing  $\Omega_{ww}$ , and hence also  $\Omega$ , to have reduced rank, we can allow for such weak unit root common factors (and thus also weak cointegration within  $x_t$ ).

**Remark 8.** Even disregarding the generality of the allowable cross-section dependencies, the current DGP is still very general when compared to existing work, in the sense that the blocks of  $\Omega$  need not be diagonal. This means that not only are the units of  $v_t$  and  $w_t$  allowed to be both serially and cross-sectionally correlated in a very general fashion, but there is also nothing to prevent the units of these variables to be correlated with each other. The types of endogeneity that can be permitted here is therefore very general indeed.

### 3 The test statistics

Let us denote by  $p$  the number of units for which  $y_{i,t}$  can be predicted using  $x_{i,t-1}$ , that is,  $p$  is the number of units for which  $\beta_i \neq 0$ . The purpose of this paper is to make inference regarding  $p$ . Let us therefore denote by  $0 = p_1 < \dots < p_K < N$  a set of  $K$  user-defined numbers, representing the number predictable of units to be considered in the testing. Let  $H_0(p_k)$  denote the null hypothesis that  $p = p_k$ , where  $k = 1, \dots, K$ , and let  $H_1(p_{k+1})$  denote the alternative hypothesis that  $p \geq p_{k+1}$ . The test statistic for testing  $H_0(p_k)$  versus  $H_1(p_{k+1})$  is henceforth going to be written in a general notation as  $\tau(p_k, p_{k+1})$ , suggesting that potentially there is a dependence on both  $p_k$  and  $p_{k+1}$ .

#### 3.1 Pooled tests for testing $p = 0$ versus $p \geq 1$

In this subsection we consider the relatively simple problem of testing  $H_0(0)$  versus  $H_1(1)$ , that is, the null hypothesis of no predictability ( $p = 0$ ) is tested versus the alternative that there is at least one predictable unit ( $p \geq 1$ ). The reason for considering this testing problem separately is that under the null hypothesis all the units of the panel are unpredictable, which makes it possible to consider pooled test statistics in the spirit of Hjalmarsen (2010). Two such test statistics are considered; one is based on the “panel” (or “within”) principle, while the other is based on the “group mean” (or “between”) principle. The exact forms of these

test statistics are as follows:

$$\begin{aligned}\tau_P(0,1) &= \frac{\sum_{i=1}^N \sum_{t=2}^T x_{i,t-1}^d y_{i,t}^d - T \hat{\lambda}_{wv}}{\sqrt{\sum_{i=1}^N \hat{\omega}_{v,i}^2 \sum_{t=2}^T (x_{i,t-1}^d)^2}}, \\ \tau_{GM}(0,1) &= \sum_{i=1}^N \theta_i,\end{aligned}$$

where

$$\theta_i = \frac{\sum_{t=2}^T x_{i,t-1}^d y_{i,t}^d - T \hat{\lambda}_{wv,i}}{\hat{\omega}_{v,i} \sqrt{\sum_{t=2}^T (x_{i,t-1}^d)^2}},$$

with  $x_{i,t-1}^d = x_{i,t-1} - T^{-1} \sum_{s=2}^T x_{i,s-1}$  and an analogous definition of  $y_{i,t}^d$ . As for the required variance correction factors, letting  $\hat{u}_t = (\hat{v}_t', \hat{w}_t')'$ , where  $\hat{v}_t$  and  $\hat{w}_t$  are the residuals obtained by applying OLS to (1) and (2), respectively, and using  $1(A)$ ,  $K(x) = (1 - |x|)1(|x| \leq 1)$  and  $J > 0$  to denote the indicator function for the event  $A$ , the Bartlett kernel and the kernel bandwidth parameter, respectively, we have

$$\hat{\Omega} = \begin{bmatrix} \hat{\Omega}_{vv} & \hat{\Omega}_{vw} \\ \hat{\Omega}_{wv} & \hat{\Omega}_{ww} \end{bmatrix} = \hat{\Sigma} + \hat{\Lambda} + \hat{\Lambda}',$$

where

$$\begin{aligned}\hat{\Sigma} &= T^{-1} \sum_{t=2}^T \hat{u}_t \hat{u}_t', \\ \hat{\Lambda} &= \sum_{j=1}^{J-1} K(j/J) T^{-1} \sum_{t=j+1}^T \hat{u}_{t-j} \hat{u}_t' .\end{aligned}$$

In this notation, using  $\hat{\omega}_{v,i}^2$  and  $\hat{\lambda}_{wv,i}$  to denote the diagonal elements of  $\hat{\Lambda}_{wv}$  and  $\hat{\Omega}_{vv}$ , respectively, we have  $\hat{\lambda}_{wv} = \text{tr}(\hat{\Lambda}_{wv}) = \sum_{i=1}^N \hat{\lambda}_{wv,i}$ , where  $\text{tr}(A)$  denoted the trace of the matrix  $A$ .

In contrast to the other test statistics that we will consider, while there is a dependence on  $p_k = p_1$  being equal to zero (in the sense that the null hypothesis of  $\beta = 0$  has been imposed),  $\tau_P(0,1)$  and  $\tau_{GM}(0,1)$  do not really depend on the proportion of predictable units under the alternative,  $p_{k+1} = p_2$ . The reason for still writing the test statistics as a function the latter is to emphasize that in case of a rejection the appropriate conclusion is that there is at least one unit for which predictability holds.

**Remark 9.** Our test statistics might seem overly complicated in the sense that in many cases one can just as well bootstrap the normalized (with respect to the sample size) OLS estimator

of the parameter of interest.<sup>3</sup> The same is true here. However, since the variance correction in the numerator cannot be dispensed with (see Remark 10 below), we can just as well bootstrap  $t$ -statistics.

**Remark 10.** The reason for bothering with the variance correction in the numerator is that otherwise the test statistics would be divergent in the case when  $m \rightarrow \infty$  (also see Theorem 3.1 in Phillips and Magdalinos, 2009). This is quite different from the conventional local-to-unity case when  $m < \infty$ , in which the effect of this bias is “second-order”. The fact that the test statistics diverge when  $m \rightarrow \infty$  is consistent with the conventional simultaneous equations bias in stationary regressions. One exception is when  $c = 0$  (not considered here), in which case  $m$  becomes irrelevant. Hence, in this case there is no need to correct the numerator.

### 3.2 Sequential test procedure for determining $p$

The tests considered in the previous section are appropriate if one wishes to infer whether there is actually any predictability at all. The problem is that in many cases one would like to go further than just concluding that  $p > 0$  in case of a rejection, and in this section we therefore consider a sequential test that can be used to pinpoint  $p$ . In so doing, we will assume that the testing numbers,  $p_1, \dots, p_K$ , are known and that  $p$  belongs to this set; later on we discuss how to proceed in general when  $p$  can lie between test numbers.

Let us denote by  $|\theta|_{(1)} \leq \dots \leq |\theta|_{(N)}$  the order statistics associated with the absolute values of  $\theta_1, \dots, \theta_N$ . The test statistic to be used in the sequential testing, denoted  $\tau_{SQ}(p_k, p_{k+1})$ , is given by the order statistic corresponding to the alternative hypothesis to be tested;

$$\tau_{SQ}(p_k, p_{k+1}) = |\theta|_{(p_{k+1})},$$

and is appropriate for testing  $H_0(p_k)$  versus  $H_1(p_{k+1})$ . Our proposed procedure for determining  $p$  is based on repeated use of this test statistic.

#### Search algorithm.

1. Use  $\tau_{SQ}(p_1, p_2)$  to test  $H_0(p_1)$  against  $H_1(p_2)$ .

---

<sup>3</sup>For example, when testing for a unit root in  $x_{i,t}$ , rather than bootstrapping the associated  $t$ -statistic, one may bootstrap  $T(\hat{\rho}_i - 1)$ , where  $\hat{\rho}_i$  is the OLS estimator of  $\rho_i$  in the  $i$ -th equation of (2).

2. If  $H_0(p_1)$  is not rejected, set  $\hat{p} = p_1$ , whereas if  $H_0(p_1)$  is rejected, use  $\tau_{SQ}(q_2, q_3)$  to test  $H_0(p_2)$  against  $H_1(p_3)$ .
3. Keep testing until  $H_0(p_k)$  cannot be rejected anymore, and set  $\hat{p} = p_k$ . If all null hypotheses up until and including  $H_0(p_K)$  are rejected, set  $\hat{p} = N$ .

**Remark 11.** The reason for using  $|\theta_{(p_{k+1})}|$  as a test statistic and not  $|\theta_{(p_k)}|$ , which might seem like a more natural choice, is that it would not test the correct set of hypotheses. For example, suppose that we are at the first stage of the search algorithm. In this case, it is quite clear that in testing  $H_0(p_1)$  against  $H_1(p_2)$ , since  $\theta_{(p_1)} = \theta_{(0)}$  is undefined, one has to take  $|\theta_{(p_2)}|$  as a test statistic.

**Remark 12.** The asymptotic theory for order statistics is known to be difficult, especially when the statistics are dependent, as in our case. The bootstrap is ideal for situations like this, and it will be used also in the present paper.

The above search algorithm is based on the assumption that the researcher knows beforehand which numbers of predictable units to test, which of course need not be the case in practice. In fact, the case with known numbers is more likely to be the exception rather than the rule. The perhaps most natural approach is to simply add the units sequentially one-by-one, which amounts to setting  $p_k = (k - 1)$  with  $k = 1, \dots, N$ . This approach has the advantage that all possible numbers are tested. The drawback is that it is likely to suffer from low power, especially when  $N$  is “large”. The reason is that when the units are added one-by-one the information used when testing one number against another is relatively small when compared to a procedure that uses wider spacing between numbers.

Fortunately, the units does not have to be added one-by-one. In fact, any pre-specified value of  $p_k$  will do. The drawback of using numbers that are further apart is that if  $p$  lies between these values, the method will be unable to detect it. For example, if the numbers to be tested are 10, 20 and 30, the method will obviously not be able to detect if 15 or 25 of the units are predictable. In other words, if  $\hat{p} = p_k$ , unless the units are added one-by-one, then this should not be taken as evidence of  $p$  being equal to  $p_k$ , but rather that  $p$  is somewhere in  $(p_{k-1}, p_{k+1})$

### 3.3 The bootstrap

For the purpose of the bootstrap algorithm, it is useful to define the following “corrected” estimators:

$$\begin{aligned}\tilde{\beta}_i &= \frac{\sum_{t=2}^T x_{i,t-1}^d y_{i,t}^d - T \hat{\lambda}_{vw,i}}{\sum_{t=2}^T (x_{i,t-1}^d)^2}, \\ \tilde{\rho}_i &= \frac{\sum_{t=2}^T x_{i,t-1}^d x_{i,t}^d - T \hat{\lambda}_{vw,i}}{\sum_{t=2}^T (x_{i,t-1}^d)^2},\end{aligned}$$

where  $\hat{\lambda}_{vw,i}$  is the  $i$ -th diagonal element of  $\hat{\Lambda}_{vw}$ . According to Lemma 1 below,  $\tilde{\beta}_i$  and  $\tilde{\rho}_i$  are  $T/\sqrt{m}$ -consistent for  $\beta_i$  and  $\rho_i$ , respectively, which is just enough to ensure validity of the bootstrap.<sup>4</sup> The next algorithm describes how  $\tilde{\beta}_i$  and  $\tilde{\rho}_i$  are used in testing  $H_0(p_k)$  versus  $H_1(p_{k+1})$ .

#### Bootstrap algorithm.

1. Let  $z_t = (x_t', y_t')'$ . Obtain  $z_t^d = (y_t^{d'}, x_t^{d'})' = z_t - T^{-1} \sum_{s=1}^T z_s$ .
2. Let  $\tilde{\gamma} = (\tilde{\beta}', \tilde{\rho}')'$ , where  $\tilde{\beta} = \text{diag}(\tilde{\beta}_1, \dots, \tilde{\beta}_N)$ ,  $\tilde{\rho} = \text{diag}(\tilde{\rho}_1, \dots, \tilde{\rho}_N)'$ ,  $z_0 = 0$ , and calculate  $\tilde{u}_t = (\tilde{v}_t', \tilde{w}_t')' = z_t^d - \tilde{\gamma} x_{t-1}^d$  for  $t = 2, \dots, T$ .
3. Choose a block length  $\ell$ . Draw  $\mathcal{I}_1, \dots, \mathcal{I}_k$  iid from the uniform distribution on  $\{1, 2, \dots, T - \ell\}$ , where  $k = \lceil (T - 1)/\ell \rceil$  is the number of blocks.
4. Construct  $u_t^* = (v_t^{*'}, w_t^{*'})' = \tilde{u}_{\mathcal{I}_{k_t} + s_t} - (T - \ell)^{-1} \sum_{\tau=1}^{T-\ell} \tilde{u}_{\tau + s_t}$ , where  $t = 2, \dots, T$ ,  $k_t = \lceil t/\ell \rceil$  and  $s_t = t - (k_t - 1)\ell$ .
5. Let  $y_t^* = v_t^*$  and  $x_t^* = \tilde{\rho} x_{t-1}^* + w_t^*$  for  $t = 2, \dots, T$  with  $x_1^* = x_1^d$  and  $y_1^* = y_1^d$ .
6. Let  $\hat{u}_t^* = (\hat{v}_t^{*'}, \hat{w}_t^{*'})'$ , where  $\hat{v}_t^*$  and  $\hat{w}_t^*$  are the residuals obtained by applying OLS to the bootstrap versions of (1) and (2), respectively. The bootstrap versions of  $\hat{\omega}_{v,i}^2$  and  $\hat{\lambda}_{vw,i}$ , denoted  $\hat{\omega}_{v,i}^{*2}$  and  $\hat{\lambda}_{vw,i}^*$ , respectively, are as before but with  $\hat{\Omega}$  and  $\hat{\Lambda}$  replaced by

$$\begin{aligned}\hat{\Omega}^* &= \frac{1}{T} \left( \sum_{m=1}^{k-1} \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} \hat{u}_{(m-1)\ell+s}^* \hat{u}_{(m-1)\ell+j}^{*'} + \sum_{s=1}^{T-(k-1)\ell} \sum_{j=1}^{T-(k-1)\ell} \hat{u}_{(k-1)\ell+s}^* \hat{u}_{(k-1)\ell+j}^{*'} \right), \\ \hat{\Lambda}^* &= \frac{1}{T} \left( \sum_{m=1}^{k-1} \sum_{s=1}^{\ell} \sum_{j=1}^{s-1} \hat{u}_{(m-1)\ell+s-j}^* \hat{u}_{(m-1)\ell+s}^{*'} + \sum_{s=1}^{T-(k-1)\ell} \sum_{j=1}^{s-1} \hat{u}_{(k-1)\ell+s-j}^* \hat{u}_{(k-1)\ell+s}^{*'} \right),\end{aligned}$$

<sup>4</sup>By contrast,  $(\hat{\beta}_i - \beta_i)$  and  $(\hat{\rho}_i - \rho_i)$  are  $O_p(m/T)$ , and therefore the algorithm cannot be based on these, as this would invalidate the resulting bootstrap.

respectively.

7. Calculate

$$\theta_i^* = \frac{\sum_{t=2}^T x_{i,t-1}^{*d} y_{i,t}^{*d} - T \hat{\lambda}_{wv,i}^*}{\hat{\omega}_{v,i}^* \sqrt{\sum_{t=2}^T (x_{i,t-1}^{*d})^2}},$$

with  $x_{i,t-1}^{*d} = x_{i,t-1}^* - T^{-1} \sum_{s=2}^T x_{i,s-1}^*$  and an analogous definition of  $y_{i,t}^{*d}$ .

8. Let  $S_k = \{i : |\theta_i| \geq |\theta|_{(p_k)}\}$  be the set of units for which the null hypothesis has already been rejected in the previous steps. Obtain  $\theta_i^*$  for all  $i \in S_k^c$ , where  $S_k^c$  is the complement of  $S_k$ . The bootstrap test statistic is given by

$$\tau_{SQ}^*(p_k, p_{k+1}) = \theta_{(p_{k+1}-p_k:S_k^c)}^*,$$

that is,  $\tau_{SQ}^*(p_k, p_{k+1})$  is the  $(p_{k+1} - p_k)$ -th smallest value of  $\theta_i^* \in S_k^c$ .

9. Repeat steps 3–8  $B$  times, and select the bootstrap critical value,  $c_\alpha^*(p_k, p_{k+1})$  say, as the  $\alpha$ -quantile of the ordered bootstrap statistics.

To perform  $\tau_P(0, 1)$  and/or  $\tau_{GM}(0, 1)$ , simply replace  $\tau_{SQ}^*(p_j, p_{j+1})$  in step 8 by  $\tau_P^*(0, 1)$  and/or  $\tau_{GM}^*(0, 1)$ , the statistics based on the bootstrap sample.

**Remark 13.** The time-specific mean correction employed in step 4 ensures that the block bootstrap sample always has zero mean. This demeaning procedure has better asymptotic properties than other procedures, and actually simplifies the proofs. It is not necessary, though. Hence, one can also demean in the usual (non-time-specific) way, as is done in, for example, Paparoditis and Politis (2003) and Palm et al. (2011).

**Remark 14.** The bootstrap variance correction explicitly takes into account the known block-wise structure of the bootstrap process. The effects of the method of studentization of block bootstrap statistics in a stationary setting have been extensively researched (see, for example, Härdle et al., 2003, Section 3). Götze and Künsch (1996) find that the type of correction used in step 6 is best in terms of refinements, which is also why we use it here.<sup>5</sup> It is, however, by no means the only correction possible. In fact, Gonçalves and Vogelsang (2011) argue, using fixed- $b$  asymptotics, that the use of the sample variance estimator in the bootstrap

---

<sup>5</sup>One disadvantage of using the correction in step 6 is that it does not guarantee that  $\hat{\Omega}^*$  and  $\hat{\Lambda}^*$  are positive definite, which, if required, can be remedied by weighting with a lag window.

ensures that the (non-negligible) effects of the chosen kernel and bandwidth can be successfully mimicked by the bootstrap, resulting in a better performance. In our simulations (see Section 5), however, we find that the bootstrap tests based on the step 6 variance estimator perform much better in finite samples than when using the sample variance estimator, which is in agreement with the results of Götze and Künsch (1996).

**Remark 15.** The selection of the block length  $\ell$  remains an open issue. It is discussed extensively in Palm et al. (2011), where it found that so-called “Warp-speed calibration” performs reasonably well. However, that method depends on the particular test statistic used, which makes it more difficult to apply in a sequential context. But since the data are the same in each step of the procedure, it seems natural to assume that also the block length remains fixed in each step. A reasonable approach in practice is therefore to apply Warp-speed calibration in the first step, and to use the resulting block length throughout. An alternative approach is to follow, for example, Gonçalves (2011) and Moon and Perron (2010), and use bandwidth selection techniques for selecting  $\ell$ .

## 4 Asymptotic distributions

Constrained by the current state of affairs of the bootstrap theory (see, for example, Palm et al., 2011), all the results reported in this section are based on keeping  $N$  fixed and sending  $T$  (and also  $m$ ) to infinity, which means that in practice what matters for accuracy is that  $T$  is “large enough”. This is different from the usual panel asymptotics, in which both  $N$  and  $T$  are passed to infinity. The downside of having  $N$  fixed is that when it comes to the effect of an increase in this quantity the results provided here are silent. The obvious advantage is that in practice  $N$  is always finite. Results developed under the assumption that  $N \rightarrow \infty$  are therefore more prone to small-sample bias than results based on having  $N$  fixed. Then there is also the problem of how to interpret the sequential test when the number of cross-section units are growing. Note in particular how the unit-by-unit test, which is perhaps the most appealing one from an applied point of view, would be rendered invalid, as asymptotically when  $N \rightarrow \infty$ ,  $p_k = (k - 1)$  is indistinguishable from  $p_{k+1} = k$ .

#### 4.1 The sample statistics

We begin by reporting the asymptotic distributions of the test statistics when applied to the sample data. The results are summarized in Theorem 1.

**Theorem 1.** *Suppose that Assumptions 1 and 2 hold,  $m = o(T^{1/2-1/\kappa})$  and  $J = o(m^{-1}\sqrt{T})$ . Then the following hold as  $m, J, T \rightarrow \infty$ :*

(i) Under  $H_0(0)$ ,

$$\begin{aligned}\tau_{GM}(0, 1) &\rightarrow_d \sum_{i=1}^N X_i, \\ \tau_P(0, 1) &\rightarrow_d \frac{\sum_{i=1}^N Y_i}{\sqrt{\text{tr}(\Omega_{vv} \odot C \odot \Omega_{ww})}},\end{aligned}$$

where  $Y = (Y_1, \dots, Y_N)' =_d (C \odot \Omega_{ww} \odot \Omega_{vv})^{1/2}Z$ , with  $Z \sim N(0, I_N)$ ,  $\rightarrow_d$  and  $=_d$  signify equality in distribution and convergence in distribution, respectively,  $\odot$  is the Hadamard product,  $C$  is a symmetric  $N \times N$  matrix with typical element  $[C]_{ij} = -1/(c_i + c_j)$ , and  $X_i = \sqrt{-2c_i}Y_i/\omega_{v,i}\omega_{w,i}$ .

(ii) Under  $H_1(1)$ ,  $\tau_{GM}(0, 1), \tau_P(0, 1) = O_p(\sqrt{m})$ .

(iii) Under  $H_0(0)$ ,

$$\tau_{SQ}(0, p) \rightarrow_d X_{(p)}.$$

(iv) Under  $H_1(q)$  with  $q > p$ ,  $\tau_{SQ}(p_k, p_{k+1}) = O_p(\sqrt{m})$ .

**Remark 16.** The requirement that  $m = o(T^{1/2-1/\kappa})$  is the same as in Park (2006). The required expansion rate of  $J$ , which is stricter than the usual  $o(\sqrt{T})$  rate (see, for example, Andrews, 1991), can be explained in the following way. As the statistic diverges with  $m$ , the convergence rate of the estimators of  $\beta$  and  $\rho$  is reduced (in the stationary case they would even become inconsistent). Therefore,  $\tilde{u}_t$  is a poor estimator of  $u_t$ , and as a result  $\hat{\Omega}$  and  $\hat{\Lambda}$  are poor estimators of  $\Omega$  and  $\Lambda$ , respectively. To compensate,  $J$  must be set as a decreasing function of  $m$ . Note also that when  $m$  increases relatively slowly, the estimation accuracy improves and the required bandwidth expansion rate becomes less restrictive. For example, if  $m \sim \ln(T)$ , then  $o(m^{-1}\sqrt{T})$  approaches the standard  $o(\sqrt{T})$  rate.



**Remark 17.** The corrections applied to the numerators and denominators of the test statistics ensure that the unit-specific nuisance parameters are eliminated, but not those arising from the cross-sectional dependence. In the time series case where  $N = 1$ , it is easy to see that  $X = \sqrt{-2c}Y/\omega_v\omega_w = \sqrt{-2c}/\omega_v\omega_w(\omega_v\omega_w Z/\sqrt{-2c}) = Z$ . Therefore,  $\theta_i$  has a limiting  $N(0, 1)$  distribution for each  $i = 1, \dots, N$ . However, if  $N > 1$  the off-diagonal elements of the covariance matrix of  $X$  are not equal to zero, but contain nuisance parameters derived from the dependence across units, which has as consequence that the panel statistics are not asymptotically pivotal; the only exception is the unrealistic situation where there is no cross-sectional dependence.

## 4.2 The bootstrap statistics

**Theorem 2.** Suppose that Assumptions 1 and 2 hold,  $m = o(T^{1/2-1/\kappa})$  and  $\ell = o(m^{-1}\sqrt{T})$ . Then the following hold as  $m, \ell, T \rightarrow \infty$ :

(i) Under  $H_0(0)$  and  $H_1(q)$  for any  $0 < q \leq N$ ,

$$\begin{aligned}\tau_{GM}^*(0, 1) &\rightarrow_{d^*} \sum_{i=1}^N X_i \text{ in probability,} \\ \tau_P^*(0, 1) &\rightarrow_{d^*} \frac{\sum_{i=1}^N Y_i}{\sqrt{\text{tr}(\Omega_{vv} \odot C \odot \Omega_{ww})}} \text{ in probability,}\end{aligned}$$

where  $\rightarrow_{d^*}$  signifies convergence in distribution conditional on the realization of the original sample.

(ii) Under  $H_0(0)$  and  $H_1(q)$ ,

$$\tau_{SQ}^*(0, p) \rightarrow_{d^*} X_{(p)} \text{ in probability.}$$

Theorem 2 shows the asymptotic validity not only of the pooled bootstrap statistics, but also of the sequential bootstrap statistics in the first step. The properties of the sequential approach as a whole is given in the following corollary to Theorem 2.

**Corollary 1.** *Under the assumptions of Theorem 2,*

$$\begin{aligned}
\lim_{T \rightarrow \infty} \mathbb{P}(\hat{p} = p_k) &= 0 && \text{if } p_{k+1} \leq p, \\
\lim_{T \rightarrow \infty} \mathbb{P}(\hat{p} = p_k) &\in [0, 1] && \text{if } p_k < p < p_{k+1}, \\
\lim_{T \rightarrow \infty} \mathbb{P}(\hat{p} = p_k) &= 1 - \alpha && \text{if } p_k = p, \\
\lim_{T \rightarrow \infty} \mathbb{P}(\hat{p} = p_k) &\in [\alpha, 1] && \text{if } p_{k-1} < p < p_k, \\
\limsup_{T \rightarrow \infty} \mathbb{P}(\hat{p} = p_k) &\leq \alpha && \text{if } p_{k-1} \geq p,
\end{aligned}$$

where  $\alpha$  is the chosen significance level.

Corollary 1 says that if  $p$  is among the numbers to be tested (as when the units are added one-by-one), the sequential method is asymptotically valid in the sense that  $\lim_{T \rightarrow \infty} \mathbb{P}(p < p_{k-1}) \leq \alpha$  and  $\lim_{T \rightarrow \infty} \mathbb{P}(p > p_{k+1}) = 0$  (see, for example, Swensen, 2006, for a similar result in the context of cointegration rank testing).

Corollary 1 also gives an insight into what happens if the true proportion is in between selected numbers. Specifically, assuming that  $\hat{p} = p_k$ , we have

$$\mathbb{P}(p \in [p_{k-1}, p_{k+1}]) = 1 - \mathbb{P}(p < p_{k-1}) - \mathbb{P}(p > p_{k+1}) \geq 1 - \alpha.$$

Hence, if the units are not added one-by-one, so that there is a possibility that  $p$  lies between the numbers considered in the testing, then the finding that  $\hat{p} = p_k$  is best interpreted as providing evidence that  $p \in [p_{k-1}, p_{k+1}]$ .

**Remark 18.** It follows from Theorem 2 that the bootstrap correctly reproduces the first-order asymptotic distributions of the panel statistics, including the nuisance parameters arising because of the cross-sectional dependence. This means that the bootstrap allows for valid and pivotal inference also in the panel case, where the asymptotic distributions are unusable without some way to estimate and correct for the nuisance parameters.

In the time series case with  $N = 1$ , or in panels without cross-sectional dependence, the standard normal approximation can be used, and therefore the bootstrap is not necessary. However, also in this case it is highly recommended to use the bootstrap, as one can use the approximation orders derived in our proofs to show along the lines of Park (2006) that the bootstrap provides asymptotic refinements if the statistic of interest is asymptotically pivotal.

## 5 Monte Carlo simulations

### 5.1 Setup

In this section we investigate briefly the performance of the proposed panel tests in small samples. The DGP used for this purpose is given by a restricted version of (1) and (2) that sets  $\alpha = \delta = 0$  and  $m = T^{1-\gamma}$ , such that  $\beta = b/T^\gamma$  and  $\rho = 1 + c/T^\gamma$ . Also,

$$u_t = \lambda f_t + \varepsilon_t, \quad (5)$$

where  $\lambda = (\lambda'_v, \lambda'_w)'$ ,  $\lambda_v = (\lambda_{v,1}, \dots, \lambda_{v,N})'$  with a similar definition of  $\lambda_w$ , and  $f_t = 0.5f_{t-1} + \varepsilon_t$  with  $\varepsilon_t \sim N(0, 1)$  independent of  $\varepsilon_t \sim N(0, I_{2N})$ . As for  $b_i$ , the  $i$ -th diagonal element of  $b$ ,  $b_i \sim U[5, 15]$  for  $i = 1, \dots, p$ , and  $b_i = 0$  for  $i = p + 1, \dots, N$ . Here we set  $p = \lfloor qN \rfloor$ , where the fraction of predictable units,  $q$ , takes the values  $q = 0, 0.2, 0.5, 0.9$ . The  $i$ -th diagonal element of  $c$ ,  $c_i$ , is made a draw from  $U[-15, 0]$ .

Two values of  $\gamma$  are considered:  $\gamma = 0.9$  and  $\gamma = 1$ . Although theoretically our bootstrap methods are invalid when  $\gamma = 1$ , as a measure of robustness, we consider it in our simulations. As in finite samples a near unit root is virtually indistinguishable from a weak unit root close to it, we still expect our tests to work reasonably well in this setting. Two values of each of  $N$  and  $T$  are considered;  $N = 10, 25$  and  $T = 100, 250$ .

In our setup, the extent of serial and cross-sectional correlation in  $v_{i,t}$  ( $w_{i,t}$ ) is determined by  $\lambda_{v,i}$  ( $\lambda_{w,i}$ ), as seen by writing  $\mathbb{E}(v_{i,t}v_{j,t-h}) = \lambda_{v,i}\lambda_{v,j}\mathbb{E}(f_tf_{t-h}) = \lambda_{v,i}\lambda_{v,j}0.5^h/(1 - 0.5^2)$ . Similarly, since  $\mathbb{E}(v_{i,t}w_{i,t-h}) = \lambda_{v,i}\lambda_{w,i}0.5^h/(1 - 0.5^2)$ , the extent of endogeneity within unit  $i$  is determined by both  $\lambda_{v,i}$  and  $\lambda_{w,i}$ . Three factor loading cases are considered:

1.  $\lambda_v = \lambda_w = 0$ :  $v_{i,t}$  and  $w_{i,t}$  are independent, both cross-sectionally and across time, and there is no endogeneity.
2.  $\lambda_{v,i} \sim U[-1, 3]$ ,  $\lambda_w = 0$ : While  $w_{i,t}$  is serially and cross-sectionally uncorrelated,  $v_{i,t}$  is not. No endogeneity is present.
3.  $\lambda_{v,i}, \lambda_{w,i} \sim U[-1, 3]$ :  $v_{i,t}$  and  $w_{i,t}$  are serially and cross-sectionally uncorrelated, and also correlated with each other (endogeneity).

In our simulations we consider  $\tau_P$ ,  $\tau_{GM}$  and  $\tau_{SQ}$ . The sequential procedure is based on setting  $p_k = k - 1$  for  $k = 1, \dots, K$ , that is, the units are added one-by-one. The block length and bandwidth are set equal to  $\ell = J = \lfloor 1.75T^{1/3} \rfloor$ , a value that was also used by Palm et al.

(2011).<sup>6</sup> All other implementation issues, including kernel and bootstrap variance estimator, are dealt with as explained in Section 3. All results are based on 1,000 simulations and 199 bootstrap replications. The significance level is set to 5%.

## 5.2 Results

The empirical rejection frequencies for  $\tau_P$  and  $\tau_{GM}$  are reported in Table 1. For  $q = 0$  these values represent size, for  $q > 0$  they represent power. It can be seen that for both panel tests size distortions are minimal. This is not only the case for the valid  $\gamma = 0.9$  setting, but also for  $\gamma = 1$ . Moreover, the tests continue to do well if endogeneity within units (case 3) is present. Although there is some oversize in this case, it is fairly minor.

As one would expect, the power of the tests increases with  $q$  and  $N$ . Moreover, for  $\gamma = 0.9$  it can also be seen that power increases with  $T$ , which is a consequence of part (ii) of Theorem 1. For  $\gamma = 1$  this does not generally seem to be the case, which is in line with theory. On average, it seems as that  $\tau_P$  is somewhat more powerful than  $\tau_{GM}$ , though for most cases considered the differences are small.

INSERT TABLE 1 ABOUT HERE

Results for the sequential tests are reported in Table 2. Here we report the average proportion of units incorrectly classified as predictable (ICP) and the average proportion of units correctly classified as predictable (CP). Here ICP can be loosely interpreted as size, while CP can be loosely interpreted as power. While for sequential/multiple testing type approaches there are a variety of performance measures (see, for example, Smeekes, 2011), for expositional simplicity we only report these two measures.

In general the sequential test results are very similar to the panel test results. Size distortions in the individual tests are small, resulting in a very small average proportion of units incorrectly classified as predictable. The ability of the sequential method to find the predictable units improves with  $q$  and with  $T$ . However, it deteriorates significantly as  $N$  increases. This is as expected from the discussion in Section 3.2, as the unit-by-unit approach does not properly utilize the cross-sectional dimension, making it more conservative for larger  $N$ . A sequential approach with  $p_k - p_{k-1} > 1$  will tend to find more predictable

---

<sup>6</sup>For the variance estimation, we also considered automatic bandwidth selection and pre-whitening, but this did not change our results much.

units, but then this will also increase the risk of some units being incorrectly classified as predictable (see Smeekes, 2011, for a more detailed discussion).

INSERT TABLE 2 ABOUT HERE

We also performed simulations in the pure time series case (with  $N = 1$ ) to investigate the performance of  $\theta_i$ , the underlying test, in more detail. In general the conclusions from those simulations match those of the panel setting reported here. One additional conclusion from the time series results that is obscured in the panel setup, is that the value of  $c$  has a significant impact on the performance of the test, in particular for small  $T$ . With  $c$  large and  $T$  relatively small, the departure from unity is relatively large, and therefore the problem of endogeneity becomes larger. As a result the variance correction does not work as well as when the deviation from unit is small, resulting in relatively large size distortions in the presence of endogeneity, though these disappear with increasing (decreasing) values of  $T$  ( $c$ ). The value of  $c$  also has an impact on power, as the larger  $c$  is, the lower power becomes. While this might not be obvious at first, it actually follows from our theoretical results, see the result (A.26) in particular. A final conclusion from our time series simulations is that, even though the bootstrap is invalid in the exact unit root case ( $c = 0$ ), in simulations it still works reasonably well even if  $c_1 = \dots = c_N = 0$  with size generally remaining well below 15%. Hence, if some units in the panel setup are close to or at a unit root, it should not affect the overall performance of the panel tests too much, a conclusion that is confirmed by our panel data results. The full set of results of the time series case are available upon request.

## 6 Concluding remarks

The difficulty of predicting stock returns using time series data, typically for the US, has recently motivated researchers to consider panel data as a means to increase the power of conventional (time series) tests. Indeed, since the predictable component of stock returns is bound to be small, if indeed one does exist, there seems to be little chance of reaching a decisive conclusion based on US data alone. Unfortunately, the few panel data tests that do exist are not only based on restrictive assumptions, but are also rather uninformative in the sense that they cannot be used to identify the units for which returns can be predicted. In the present paper we take this as our starting point to develop a block bootstrap algorithm that can be used to infer panel predictive regressions under very general conditions. Two tests

based on this bootstrap are proposed. One is suitable when testing the null hypothesis of no predictability versus the general alternative, while the other can be used to identify exactly the units for which predictability holds. The asymptotic validity of the tests is proven and verified in samples using Monte Carlo simulations.

## References

- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies* 20, 651–707.
- Bai, J. and S. Ng (2010). Panel unit root tests with cross-section dependence: A further investigation. *Econometric Theory* 26, 1088–1114.
- Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics* 16, 1709–1722.
- Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.
- Chang, Y., J. Y. Park, and K. Song (2006). Bootstrapping cointegrating regressions. *Journal of Econometrics* 133, 703–739.
- Driesprong, G., B. Jacobsen, and B. Maat (2008). Striking oil: another puzzle? *Journal of Financial Economics* 89, 307–327.
- Giraitis, L. and P. C. B. Phillips (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis* 27, 51–60.
- Gonçalves, S. (2011). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory* 27, 1048–1082.
- Gonçalves, S. and T. J. Vogelsang (2011). Block bootstrap HAC robust tests: the sophistication of the naive bootstrap. *Econometric Theory* 27, 745–791.
- Götze, F. and H. R. Künsch (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914–1033.
- Hannan, E. J. (1970). *Multiple Time Series*. New York: Wiley.
- Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 498–500.

- Härdle, W., J. L. Horowitz, and J.-P. Kreiss (2003). Bootstrap methods for time series. *International Statistical Review* 71, 435–459.
- Hjalmarsson, E. (2008). The stambaugh bias in panel predictive regressions. *Finance Research Letters* 5, 45–58.
- Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis* 45, 49–80.
- Kauppi, H. (2001). Panel data limit theory and asymptotic analysis of a panel regression with near integrated regressors. In B. H. Baltagi, T. B. Fomby, and R. C. Hill (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Volume 15 of *Advances in Econometrics*, pp. 239–274. Emerald Group Publishing Limited.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Lanne, M. (2002). Testing the predictability of stock returns. *The Review of Economics and Statistics* 84, 407–415.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics* 74, 209–235.
- Moon, H. R. and B. Perron (2010). Beyond panel unit root tests: Using multiple testing to determine the non stationarity properties of individual series in a panel. Department of Economics working paper 2010-04, University of Montreal.
- Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2012). Forecasting the equity risk premium: the role of technical indicators. Federal Reserve Bank of St. Louis Working Paper 2010-008E, Federal Reserve Bank of St. Louis.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2011). Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics* 163, 85–104.
- Paparoditis, E. and D. N. Politis (2003). Residual-based block bootstrap for unit root testing. *Econometrica* 71, 813–855.
- Park, J. Y. (2003). Weak unit roots. Working Paper 2003–17, Department of Economics, Rice University.



- Park, J. Y. (2006). A bootstrap theory for weakly integrated processes. *Journal of Econometrics* 133, 639–672.
- Phillips, P. C. B. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
- Phillips, P. C. B. and T. Magdalinos (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory* 25, 482–526.
- Phillips, P. C. B., T. Magdalinos, and L. Giraitis (2010). Smoothing local-to-moderate unit root theory. *Journal of Econometrics* 158, 274–279.
- Polk, C., S. Thompson, and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics* 81, 101–141.
- Rapach, D. E., J. K. Strauss, and G. Zhou (2012). International stock return predictability: What is the role of the United States? *Journal of Finance*, forthcoming.
- Rapach, D. E. and G. Zhou (2012). Forecasting stock returns. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, pp. forthcoming. Elsevier.
- Smeeke, S. (2011). Bootstrap sequential tests to determine the stationary units in a panel. METEOR Research Memorandum 11/003, Maastricht University.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics* 54, 375–421.
- Stock, J. H. (1997). Cointegration, long-run comovements, and long-horizon forecasting. In D. M. Kreps and K. F. Wallis (Eds.), *Advances in Economics and Econometrics: Theory and Applications: Seventh World Congress*, Volume 3, Chapter 2, pp. 34–60. Cambridge: Cambridge University Press.
- Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* 74, 1699–1714.
- Westerlund, J. and P. K. Narayan (2012). Does the choice of estimator matter when forecasting returns? *Journal of Banking and Finance* 36, 2632–2640.
- Wolf, M. (2000). Stock returns and dividend yields revisited: A new way to look at an old problem. *Journal of Business and Economic Statistics* 18, 18–30.

## A Appendix: Proofs for the sample statistics

**Lemma A.1.** *Under Assumptions 1 and 2, uniformly in  $m$ ,*

$$\begin{aligned} \frac{\sqrt{m}}{T} \left( \sum_{t=2}^T x_{t-1}^d (v_t^d)' - T \Lambda_{wv} \right) &= {}_d \sqrt{m} \mathcal{A}_m + O_p(\sqrt{m} T^{-1/2+1/\kappa}) + O_p(m^{3/2} T^{-1}), \\ \frac{m}{T^2} \sum_{t=2}^T x_{t-1}^d (x_{t-1}^d)' &= {}_d m \mathcal{B}_m + O_p(\sqrt{m} T^{-1/2+1/\kappa}) + O_p(m^{3/2} T^{-1}), \end{aligned}$$

where

$$\begin{aligned} \mathcal{A}_m &= \int_0^1 J_{cm}^d(s) dB_v(s)', \\ \mathcal{B}_m &= \int_0^1 J_{cm}^d(s) J_{cm}^d(s)' ds, \end{aligned}$$

where  $=_d$  signifies equality in distribution,  $J_c^d(s) = J_c(s) - \int_0^1 J_c(r) dr$ ,  $J_c(s) = \int_0^s \exp(c(s-r)) dB_w(r)$ ,  $B(s) = [B_w(s)', B_v(s)']' = \Omega^{1/2} W(s)$  and  $W(s)$  is a  $2N$ -vector standard Brownian motion.

### Proof of Lemma A.1.

Consider the first result. Write  $\Psi(z) = (\Psi_v(z)', \Psi_w(z)')'$ , where  $\Psi_v(z)$  and  $\Psi_w(z)$  have dimension  $N \times 2N$ . By the Beveridge–Nelson (BN) decomposition,  $\Psi(z) = \Psi(1) - (1-z)\Psi^*(z)$ , where  $\Psi^*(z) = \sum_{j=0}^{\infty} \Psi_j^* z^j$  with  $\Psi_j^* = \sum_{k=j+1}^{\infty} \Psi_k$ . Letting  $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{v,t}', \tilde{\varepsilon}_{w,t}')' = \Psi^*(L)\varepsilon_t$ , it follows that

$$u_t = \Psi(z)\varepsilon_t = \Psi(1)\varepsilon_t - \Delta \tilde{\varepsilon}_t,$$

which in turn implies

$$\frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d v_t' = \frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d \varepsilon_t' \Psi_v(1)' - \frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d \Delta \tilde{\varepsilon}_{v,t}'. \quad (\text{A.1})$$

Consider the first term on the right-hand side, where, ignoring the demeaning for the moment,

$$\begin{aligned} \frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d \varepsilon_t' \Psi_v(1)' &= \Psi_w(1) \frac{\sqrt{m}}{T} \sum_{t=2}^T \tilde{x}_{t-1}^d \varepsilon_t' \Psi_v(1)' \\ &\quad + \frac{\sqrt{m}}{T} \sum_{t=2}^T (x_{t-1} - \Psi_w(1) \tilde{x}_{t-1}) \varepsilon_t' \Psi_v(1)', \end{aligned} \quad (\text{A.2})$$

with  $\varepsilon_t = (\varepsilon'_{v,t}, \varepsilon'_{w,t})'$  and  $\tilde{x}_t = \sum_{s=1}^t \rho^{t-s} \varepsilon_{w,t}$ . By using  $x_0 = 0$ , the BN decomposition of  $\Psi(z)$ , summation by parts, and then  $\Delta \rho^{t-s} = \rho^{t-s} - \rho^{t-s-1} = (I_N - \rho) \rho^{t-s} = -T^{-1} m c \rho^{t-s}$ ,

$$\begin{aligned} x_t &= \sum_{s=1}^t \rho^{t-s} w_s = \sum_{s=1}^t \rho^{t-s} w_s = \Psi_w(1) \tilde{x}_t + \sum_{s=1}^t \rho^{t-s} \Delta \tilde{w}_{w,s} \\ &= \Psi_w(1) \tilde{x}_t + \tilde{\varepsilon}_{w,t} - \sum_{s=1}^t \Delta \rho^{t-s} \tilde{\varepsilon}_{w,s-1} \\ &= \Psi_w(1) \tilde{x}_t + \tilde{\varepsilon}_{w,t} + \frac{mc}{T} \sum_{s=1}^t \rho^{t-s} \tilde{\varepsilon}_{w,s-1}, \end{aligned}$$

from which it follows that

$$\begin{aligned} &\frac{\sqrt{m}}{T} \sum_{t=2}^T (x_{t-1} - \Psi_w(1) \tilde{x}_{t-1}) \varepsilon'_t \Psi_v(1)' \\ &= \frac{\sqrt{m}}{T} \sum_{t=2}^T (\tilde{\varepsilon}_{w,t-1} + mcT^{-1} \sum_{s=1}^{t-1} \rho^{t-1-s} \tilde{\varepsilon}_{w,s-1}) \varepsilon'_t \Psi_v(1)'. \end{aligned} \quad (\text{A.3})$$

Let us denote by  $\mathcal{F}_t$  denote the smallest sigma-field containing all past information up to time  $t$  on  $\varepsilon_t$ , and define  $\mathbb{E}_t X = \mathbb{E}(X|\mathcal{F}_t)$ . As usual,  $\otimes$  denotes the Kronecker product and  $\text{vec}(A)$  is the vectorization of  $A$ . Clearly,  $\mathbb{E} \tilde{\varepsilon}_{w,t-1} \varepsilon'_t = \mathbb{E}(\tilde{\varepsilon}_{w,t-1} \mathbb{E}_{t-1} \varepsilon'_t) = 0$ , and it also not difficult to show that, with  $\text{vec}(AB') = B \otimes A$  and  $(A \otimes B)(C \otimes D) = AC \otimes BD$ ,

$$\begin{aligned} &\mathbb{E} \left[ \text{vec} \left( \frac{\sqrt{m}}{T} \sum_{t=2}^T \tilde{\varepsilon}_{w,t-1} \varepsilon'_t \right) \text{vec} \left( \frac{\sqrt{m}}{T} \sum_{s=2}^T \tilde{\varepsilon}_{w,s-1} \varepsilon'_s \right)' \right] \\ &= \frac{m}{T^2} \sum_{t=2}^T \mathbb{E}(\mathbb{E}_{t-1} \varepsilon_t \varepsilon'_t \otimes \tilde{\varepsilon}_{w,t-1} \tilde{\varepsilon}'_{w,t-1}) = \frac{m}{T^2} \sum_{t=2}^T (\Sigma_{\varepsilon\varepsilon} \otimes \mathbb{E} \tilde{\varepsilon}_{w,t-1} \tilde{\varepsilon}'_{w,t-1}) = O(T^{-1}m), \end{aligned}$$

suggesting that

$$\left\| \frac{\sqrt{m}}{T} \sum_{t=2}^T \tilde{\varepsilon}_{w,t-1} \varepsilon'_t \right\| = O(T^{-1/2} \sqrt{m}).$$

It follows from Lemma 4.2 in Park (2006) that  $\|x_t\| = O_p(\sqrt{T}m^{-1/2})$ , which, together with  $\mathbb{E} x_{t-1} \varepsilon_t = 0$ , implies  $\|\sum_{t=2}^T x_{t-1} \varepsilon'_t\| = O_p(Tm^{-1/2})$ . Asymptotically,  $\sum_{s=1}^{t-1} \rho^{t-1-s} \tilde{\varepsilon}_{w,s-1}$  behaves like  $x_t$ , and in fact satisfies an invariance principle (when suitably normalized by  $m$  and  $T$ ). Hence, in analogy to  $\sum_{t=2}^T x_{t-1} \varepsilon'_t$ ,

$$\left\| \frac{m^{3/2}}{T^2} \sum_{t=2}^T \sum_{s=1}^{t-1} \rho^{t-1-s} \tilde{\varepsilon}_{w,s-1} \varepsilon'_t \right\| = O(T^{-1}m).$$

Consequently,

$$\frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1} \varepsilon'_t \Psi_v(1)' = \Psi_w(1) \frac{\sqrt{m}}{T} \sum_{t=2}^T \tilde{x}_{t-1} \varepsilon'_t \Psi_v(1)' + O(T^{-1/2} \sqrt{m}). \quad (\text{A.4})$$

The demeaning in  $x_t^d$  does not affect the order of the remainder. Hence, using  $\tilde{x}_t^d$  to denote the demeaned version of  $\tilde{x}_t$ ,

$$\frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d \varepsilon_t' \Psi_v(1)' = \Psi_w(1) \frac{\sqrt{m}}{T} \sum_{t=2}^T \tilde{x}_{t-1}^d \varepsilon_t' \Psi_v(1)' + O(T^{-1/2} \sqrt{m}). \quad (\text{A.5})$$

Let  $W(s) = [W_w(s)', W_v(s)']'$ ,  $J_c(s) = \int_0^1 \exp(c(s-r)) dW_w(r)$  and  $J_c^d(s) = J_c(s) - \int_0^1 J_c(r) dr$ . Then, by Lemma 3.1 of Park (2006),

$$\frac{\sqrt{m}}{T} \sum_{t=2}^T \tilde{x}_{t-1}^d \varepsilon_t' =_d \sqrt{m} \int_0^1 J_{cm}^d(s) dW_v(s)' + O_p(\sqrt{m} T^{-1/2+1/\kappa}) + O_p(m^{3/2} T^{-1}). \quad (\text{A.6})$$

Hence, since  $O_p(m^{3/2} T^{-1}) > O(T^{-1/2} \sqrt{m})$ , defining  $\mathcal{A}_m = \int_0^1 J_{cm}^d(s) dB_v(s)'$ , by the continuous mapping theorem,

$$\frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d \varepsilon_t' \Psi_v(1)' =_d \sqrt{m} \mathcal{A}_m + O_p(\sqrt{m} T^{-1/2+1/\kappa}) + O_p(m^{3/2} T^{-1}). \quad (\text{A.7})$$

Next, consider  $\sqrt{m} T^{-1} \sum_{t=2}^T x_{t-1}^d \Delta \tilde{\varepsilon}_{v,t}'$ . From  $\|x_t\| = O_p(\sqrt{T} m^{-1/2})$ ,

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^T x_{t-1}^d \Delta \tilde{\varepsilon}_t' &= -\frac{1}{T} \sum_{t=2}^T \Delta x_t \tilde{\varepsilon}_t' + \frac{1}{T} \sum_{t=2}^T (x_t \tilde{\varepsilon}_t' - x_{t-1} \tilde{\varepsilon}_{t-1}') - \frac{1}{T^2} \sum_{s=2}^T x_{s-1} \sum_{t=2}^T \Delta \tilde{\varepsilon}_t' \\ &= -\frac{1}{T} \sum_{t=2}^T \Delta x_t \tilde{\varepsilon}_t' + \frac{1}{T} x_T \tilde{\varepsilon}_T' - \frac{1}{T^2} \sum_{t=2}^T x_{t-1} (\tilde{\varepsilon}_T - \tilde{\varepsilon}_1)' \\ &= -\frac{1}{T} \sum_{t=2}^T \Delta x_t \tilde{\varepsilon}_t' + O_p((mT)^{-1/2}) \end{aligned}$$

and by further use of (2),  $\rho = 1 + cmT^{-1}$  and  $\sum_{t=2}^T x_{t-1} \varepsilon_t' = O_p(Tm^{-1/2})$ ,

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^T \Delta x_t \tilde{\varepsilon}_t' &= \frac{1}{T} \sum_{t=2}^T w_t \tilde{\varepsilon}_t' + c \frac{m}{T^2} \sum_{t=2}^T x_{t-1} \tilde{\varepsilon}_t' = \frac{1}{T} \sum_{t=2}^T w_t \tilde{\varepsilon}_t' + O_p(mT^{-1}) \\ &= \frac{1}{T} \sum_{t=2}^T \Psi_w(L) \varepsilon_t \tilde{\varepsilon}_t' + O_p(mT^{-1}). \end{aligned}$$

By definition,

$$\Lambda = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \Psi_j \Sigma_{\varepsilon\varepsilon} \Psi_{j+k}' = \sum_{j=0}^{\infty} \Psi_j \Sigma_{\varepsilon\varepsilon} \left( \sum_{k=1}^{\infty} \Psi_{j+k} \right)' = \sum_{j=0}^{\infty} \Psi_j \Sigma_{\varepsilon\varepsilon} \left( \sum_{k=j+1}^{\infty} \Psi_k \right)' = \sum_{j=0}^{\infty} \Psi_j \Sigma_{\varepsilon\varepsilon} \Psi_j^{*'},$$

which is identically  $\mathbb{E} u_t \tilde{\varepsilon}_t'$ . It follows that

$$\frac{1}{T} \left( \sum_{t=2}^T w_t \tilde{\varepsilon}_{v,t}' - T \Lambda_{wv} \right) = O_p(T^{-1/2}).$$

Thus, by combining the results,

$$\frac{\sqrt{m}}{T} \left( \sum_{t=2}^T x_{t-1}^d \Delta \tilde{\varepsilon}_t' - T \Lambda_{wv} \right) = O_p(\sqrt{m} T^{-1/2}) + O_p(m^{3/2} T^{-1}), \quad (\text{A.8})$$

which in turn implies

$$\begin{aligned} \frac{\sqrt{m}}{T} \left( \sum_{t=2}^T x_{t-1}^d v'_t - T \Lambda_{wv} \right) &= \frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^d \varepsilon'_t \Psi_v(1)' + O_p(\sqrt{m} T^{-1/2+1/\kappa}) \\ &\quad + O_p(m^{3/2} T^{-1}) \\ &=_d \sqrt{m} \mathcal{A}_m + O_p(\sqrt{m} T^{-1/2+1/\kappa}) + O_p(m^{3/2} T^{-1}). \end{aligned} \quad (\text{A.9})$$

It then also follows straightforwardly that

$$\begin{aligned} \frac{\sqrt{m}}{T} \left( \sum_{t=2}^T x_{t-1}^d (v_t^d)' - T \Lambda_{wv} \right) &= \frac{\sqrt{m}}{T} \left( \sum_{t=2}^T x_{t-1}^d v'_t - T \Lambda_{wv} \right) + O_p(T^{-1/2}) \\ &=_d \sqrt{m} \mathcal{A}_m + O_p(\sqrt{m} T^{-1/2+1/\kappa}) + O_p(m^{3/2} T^{-1}). \end{aligned} \quad (\text{A.10})$$

The second result follows directly from the proof of Lemma 3.1 in Park (2006).  $\blacksquare$

**Lemma A.2.** *Under the conditions of Lemma A.1, as  $m \rightarrow \infty$*

$$\begin{aligned} \text{diag}(\sqrt{m} \mathcal{A}_m) &\rightarrow_d Y, \\ m \mathcal{B}_m &\rightarrow_p C \odot \Omega_{ww}, \end{aligned}$$

where  $Y = (Y_1, \dots, Y_N)' =_d (C \odot \Omega_{ww} \odot \Omega_{vv})^{1/2} Z$  with  $Z \sim N(0, I_N)$ ,  $\rightarrow_d$  and  $\rightarrow_p$  signify convergence in distribution and probability, respectively,  $\odot$  is the Hadamard product, and the  $N \times N$  matrix  $C$  has typical element  $[C]_{ij} = -1/(c_i + c_j)$ .

**Proof of Lemma A.2.**

Consider  $\sqrt{m} \mathcal{A}_m$ . By variable substitution of  $u = ms$ ,  $dB_w(s) =_d \sqrt{m} dB_w(s/m) = \sqrt{m} dB_w(u)$ .

Thus, letting  $v = mr$ ,

$$\begin{aligned} J_{cm}(s) &= \int_0^s \exp(cm(s-r)) dB_w(r) = \frac{1}{\sqrt{m}} \int_0^u \exp(c(u-v)) \sqrt{m} dB_w(vm^{-1}) \\ &=_d \frac{1}{\sqrt{m}} \int_0^u \exp(c(u-v)) dB_w(v) = \frac{1}{\sqrt{m}} J_c(u), \end{aligned} \quad (\text{A.11})$$

from which it follows that

$$J_{cm}^d(s) =_d \frac{1}{\sqrt{m}} J_c(u) - \frac{1}{m^{3/2}} \int_0^m J_c(s) ds = \frac{1}{\sqrt{m}} J_c(u) - \frac{1}{m^{3/2}} \sum_{j=1}^m \int_{j-1}^j J_c(s) ds$$

(see also Phillips et al., 2010, page 275). By using this, and a central limit theorem for martingale difference processes,

$$\begin{aligned} \sqrt{m} \mathcal{A}_m &= \frac{1}{\sqrt{m}} \sum_{j=1}^m \int_{j-1}^j J_c(u) dB_v(u)' - \frac{1}{m} \sum_{i=1}^m \int_{i-1}^i J_c(v) dv \frac{1}{\sqrt{m}} \sum_{j=1}^m \int_{j-1}^j dB_v(u)' \\ &= \frac{1}{\sqrt{m}} \sum_{j=1}^m \int_{j-1}^j J_c(u) dB_v(u)' + O_p(m^{-1/2}), \end{aligned} \quad (\text{A.12})$$

where the remaining term is clearly mean zero. As for the covariance matrix of this term, by using the same steps as in the proof of Lemma A.1 and  $\mathbb{E} dB_v(u)dB_v(u)' = \Omega_{vv}du$ ,

$$\begin{aligned} & \mathbb{E} \left[ \text{vec} \left( \int_{j-1}^j J_c(u) dB_v(u)' \right) \text{vec} \left( \int_{j-1}^j J_c(r) dB_v(r)' \right)' \right] \\ &= \int_{j-1}^j \mathbb{E} [\mathbb{E} dB_v(u) dB_v(u)' \otimes J_c(u) J_c(u)'] = \int_{j-1}^j (\Omega_{vv} \otimes \mathbb{E} J_c(u) J_c(u)') du. \end{aligned}$$

where

$$\begin{aligned} \mathbb{E} J_c(u) J_c(u)' &= \int_0^u \int_0^u \exp(c(u-r)) \mathbb{E} dB_w(r) dB_w(v) \exp(c(u-v))' \\ &= \int_0^u \exp(c(u-r)) \Omega_{ww} \exp(c(u-r))' dr. \end{aligned}$$

Provided that  $c_i < 0$ , letting  $\omega_{w,i}^2 = [\Omega_{ww}]_{ii}$  and  $\omega_{w,ik} = [\Omega_{ww}]_{ik}$  for  $i \neq k$ , the typical element of the above expectation is given by

$$\omega_{w,ik} \int_0^u \exp((c_i + c_k)(u-r)) dr = -\frac{\omega_{w,ik}}{(c_i + c_k)} (1 - \exp((c_i + c_k)u)),$$

suggesting that the typical element of  $\int_{j-1}^j \mathbb{E} J_c(u) J_c(u)' du$  can be obtained as

$$\begin{aligned} & -\frac{\omega_{w,ik}}{(c_i + c_k)} \int_{j-1}^j (1 - \exp((c_i + c_k)u)) du \\ &= -\frac{\omega_{w,ik}}{(c_i + c_k)} \left( 1 - \frac{1}{(c_i + c_k)} \exp(j(c_i + c_k)) (1 - \exp(-(c_i + c_k))) \right) \rightarrow -\frac{\omega_{w,ik}}{(c_i + c_k)}, \end{aligned}$$

where the last result holds because  $c_i < 0$ , and therefore  $\exp(j(c_i + c_k)) \rightarrow 0$  as  $j \rightarrow \infty$ . Now, we are only interested in the covariance matrix corresponding to the diagonal elements of  $\int_{j-1}^j J_c(u) dB_v(u)'$ , which, in view of the above, can be written as  $C \odot \Omega_{ww} \odot \Omega_{vv}$ , where

$$C = - \begin{bmatrix} 1/2c_1 & \cdots & 1/(c_1 + c_N) \\ \vdots & \ddots & \vdots \\ 1/(c_N + c_1) & \cdots & 1/2c_N \end{bmatrix}.$$

Therefore, by a central limit theorem for martingale difference processes,

$$\text{diag} \left( \frac{1}{\sqrt{m}} \sum_{j=1}^m \int_{j-1}^j J_c(u) dB_v(u)' \right) \rightarrow_d (C \odot \Omega_{ww} \odot \Omega_{vv})^{1/2} Z \quad (\text{A.13})$$

as  $m \rightarrow \infty$ .

As for  $m\mathcal{B}_m$ , by using the same trick as before,

$$m\mathcal{B}_m = \frac{1}{m} \int_0^m J_c^d(u) J_c^d(u)' du = \frac{1}{m} \sum_{j=1}^m \int_{j-1}^j J_c(u) J_c(u)' du + O_p(m^{-1}), \quad (\text{A.14})$$

where, via the ergodic theorem,

$$\frac{1}{m} \sum_{j=1}^m \int_{j-1}^j J_c(u) J_c(u)' du \rightarrow_p C \odot \Omega_{ww} \quad (\text{A.15})$$

as  $m \rightarrow \infty$ . ■

**Lemma A.3.** *Under the conditions of Lemma A.1, as  $J, T \rightarrow \infty$ ,*

$$\begin{aligned} \|\hat{\Lambda} - \Lambda\| &= O_p(\sqrt{J}T^{-1/2}) + O_p(J^{3/2}T^{-1}) + O_p(\sqrt{m}JT^{-1/2}), \\ \|\hat{\Omega} - \Omega\| &= O_p(\sqrt{J}T^{-1/2}) + O_p(J^{3/2}T^{-1}) + O_p(\sqrt{m}JT^{-1/2}). \end{aligned}$$

**Proof of Lemma A.3.**

Consider  $\hat{\Sigma} = T^{-1} \sum_{t=2}^T \hat{u}_t \hat{u}_t'$ , where  $\hat{u}_t = (\hat{w}_t', \hat{v}_t')'$ ,  $\hat{w}_t = x_t^d - \hat{\rho}x_{t-1}^d$ ,  $\hat{v}_t = y_t^d - \hat{\beta}x_{t-1}^d$ ,  $\hat{\rho} = \text{diag}(\hat{\rho}_1, \dots, \hat{\rho}_N)$  and  $\hat{\beta} = \text{diag}(\hat{\beta}_1, \dots, \hat{\beta}_N)$ . From (1),

$$m^{-1/2}T(\hat{\beta}_i - \beta_i) = \frac{\sqrt{m}T^{-1}(\sum_{t=2}^T x_{i,t-1}^d v_{i,t}^d - T\Lambda'_{vw})}{mT^{-2} \sum_{t=2}^T (x_{i,t-1}^d)^2} + \frac{\sqrt{m}\Lambda'_{vw}}{mT^{-2} \sum_{t=2}^T (x_{i,t-1}^d)^2},$$

which is  $O_p(1)$  if  $m < \infty$  (Lemma A.1) and  $O_p(\sqrt{m})$  if  $m \rightarrow \infty$  (Lemma A.2). Hence,  $(\hat{\beta}_i - \beta_i) = O_p(mT^{-1})$ . This result, together with the fact that  $\hat{v}_t = v_t^d - (\hat{\beta} - \beta)x_{t-1}^d$ , imply that the upper left block of  $\hat{\Sigma}$  can be written as

$$\begin{aligned} \hat{\Sigma}_{vv} &= \frac{1}{T} \sum_{t=2}^T \hat{v}_t \hat{v}_t' = \frac{1}{T} \sum_{t=2}^T v_t^d (v_t^d)' - \frac{1}{T} \sum_{t=2}^T v_t^d (x_{t-1}^d)' (\hat{\beta} - \beta)' \\ &\quad - (\hat{\beta} - \beta) \frac{1}{T} \sum_{t=2}^T x_{t-1}^d (v_t^d)' + (\hat{\beta} - \beta) \frac{1}{T} \sum_{t=2}^T x_{t-1}^d (x_{t-1}^d)' (\hat{\beta} - \beta)' \\ &= \frac{1}{T} \sum_{t=2}^T v_t^d (v_t^d)' + O_p(mT^{-1}) = \Sigma_{vv} + \frac{1}{T} \sum_{t=2}^T (v_t v_t' - \Sigma_{vv}) + O_p(mT^{-1}) \\ &= \Sigma_{vv} + O_p(T^{-1/2}) + O_p(mT^{-1}). \end{aligned}$$

Similar results apply to the other blocks of  $\hat{\Sigma}$ . Hence,

$$\hat{\Sigma} = \Sigma + O_p(T^{-1/2}) + O_p(mT^{-1}). \quad (\text{A.16})$$

Next, consider  $\hat{\Lambda}$ . In particular, let us consider  $\hat{\Lambda}_{wv}$ , which we can expand as

$$\begin{aligned}
\hat{\Lambda}_{wv} &= \sum_{j=1}^{J-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T \hat{w}_{t-j} \hat{v}_t' \\
&= \sum_{j=1}^{M-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T w_{t-j} v_t' + \sum_{j=1}^{J-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T \hat{w}_{t-j} (\hat{v}_t - v_t)' \\
&\quad + \sum_{j=1}^{M-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T (\hat{w}_{t-j} - w_{t-j}) \hat{v}_t' \\
&\quad - \sum_{j=1}^{J-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T (\hat{w}_{t-j} - w_{t-j}) (\hat{v}_t - v_t)',
\end{aligned} \tag{A.17}$$

where the last three terms on the right-hand side are dominated by second and third terms. Consider the second term. By the Cauchy–Schwarz inequality and then  $\hat{v}_t = v_t^d - (\hat{\beta} - \beta)x_{t-1}^d = v_t^d + O_p(\sqrt{m}T^{-1/2})$ , as follows from noting that  $x_{i,t-1}^d = O_p(\sqrt{T}m^{-1/2})$  (see the proof of Lemma A.2),

$$\begin{aligned}
&\left\| \sum_{j=1}^{J-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T w_{t-j} (\hat{v}_t - v_t)' \right\| \\
&\leq \sum_{j=1}^{J-1} K(j/J) \left( \frac{1}{T} \sum_{t=j+1}^T \|w_{t-j}\|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=j+1}^T \|\hat{v}_t - v_t\|^2 \right)^{1/2} = O_p(\sqrt{m}JT^{-1/2}).
\end{aligned}$$

Thus, since the third term is of the same order,

$$\hat{\Lambda}_{wv} = \sum_{j=1}^{J-1} K(j/J) \frac{1}{T} \sum_{t=j+1}^T w_{t-j} v_t' + O_p(\sqrt{m}JT^{-1/2}) = \tilde{\Lambda}_{wv} + O_p(\sqrt{m}JT^{-1/2}), \tag{A.18}$$

with an obvious definition of  $\tilde{\Lambda}_{wv}$ . Observe that

$$\mathbb{E} \|\tilde{\Lambda}_{wv} - \Lambda_{wv}\|^2 \leq \mathbb{E} \|\tilde{\Lambda}_{wv} - \mathbb{E} \tilde{\Lambda}_{wv}\|^2 + \|\mathbb{E}(\tilde{\Lambda}_{wv} - \Lambda_{wv})\|^2 = O(JT^{-1}), \tag{A.19}$$

where we have used that, following the same lines of proof as that in Theorems 9 and 10 Hannan (1970, pages 280–283) (or Proposition 1 in Andrews, 1991),  $\mathbb{E} \|\tilde{\Lambda}_{wv} - \mathbb{E} \tilde{\Lambda}_{wv}\|^2 = O(JT^{-1})$  and  $\mathbb{E} \|\tilde{\Lambda}_{wv} - \Lambda_{wv}\| = O(J^{-1})$ . Hence,  $\mathbb{E} \|\tilde{\Lambda}_{wv} - \Lambda_{wv}\| = O(\sqrt{J}T^{-1/2})$ .

It then follows from the Markov inequality that

$$\mathbb{P}(\sqrt{J}T^{-1/2} \|\tilde{\Lambda}_{wv} - \Lambda_{wv}\| > \delta) \leq \delta^{-2} JT^{-1} \mathbb{E} \|\tilde{\Lambda}_{wv} - \Lambda_{wv}\|^2 = O(1),$$

and therefore  $\|\tilde{\Lambda}_{wv} - \Lambda_{wv}\| = O_p(\sqrt{J}T^{-1/2})$ . Consequently,

$$\begin{aligned}
\|\hat{\Lambda}_{wv} - \Lambda_{wv}\| &\leq \|\hat{\Lambda}_{wv} - \tilde{\Lambda}_{wv}\| + \|\tilde{\Lambda}_{wv} - \Lambda_{wv}\| \\
&= O_p(\sqrt{m}JT^{-1/2}) + O_p(\sqrt{J}T^{-1/2}).
\end{aligned} \tag{A.20}$$



The same is true for the other blocks of  $\hat{\Lambda}$ . Moreover, since  $O_p(\sqrt{J}T^{-1/2}) > O_p(mT^{-1})$ , we have that

$$\begin{aligned} \|\hat{\Omega} - \Omega\| &\leq \|\hat{\Sigma} - \Sigma\| + 2\|\hat{\Lambda} - \Lambda\| = O_p(\sqrt{J}T^{-1/2}) + O_p(J^{3/2}T^{-1}) \\ &\quad + O_p(\sqrt{m}JT^{-1/2}), \end{aligned} \tag{A.21}$$

and so the proof is complete.  $\blacksquare$

### Proof of Theorem 1.

From Lemmas A.1–A.3, we can deduce easily the asymptotic distribution of  $\theta_i$ . In fact, by using (1),  $\beta_i = b_i m/T$ , Lemmas A.1 and A.2, and then summation by parts,

$$\begin{aligned} \theta_i &= \frac{\sum_{t=2}^T x_{i,t-1}^d y_{i,t}^d - T\hat{\lambda}_{vw,i}}{\hat{\omega}_{v,i} \sqrt{\sum_{t=2}^T (x_{i,t-1}^d)^2}} \\ &= \frac{m^{3/2}T^{-2} \sum_{t=2}^T b_i (x_{i,t-1}^d)^2 + \sqrt{m}T^{-1} (\sum_{t=2}^T x_{i,t-1}^d v_{i,t}^d - T\hat{\lambda}_{vw,i})}{\hat{\omega}_{v,i} \sqrt{mT^{-2} \sum_{t=2}^T (x_{i,t-1}^d)^2}} \\ &\quad + \frac{\sqrt{m}(\hat{\lambda}_{vw,i} - \lambda_{vw,i})}{\hat{\omega}_{v,i} \sqrt{mT^{-2} \sum_{t=2}^T (x_{i,t-1}^d)^2}} \\ &= {}_d \mathcal{D}_{m,i} + O_p(\sqrt{m}JT^{-1/2}) + O_p(\sqrt{m}J^{3/2}T^{-1}) + O_p(mJT^{-1/2}) \\ &\quad + O_p(\sqrt{m}T^{-1/2+1/\kappa}) + O_p(m^{3/2}T^{-1}), \end{aligned} \tag{A.22}$$

where

$$\mathcal{D}_{m,i} = \frac{mb_i \mathcal{B}_{m,i} + \mathcal{A}_{m,i}}{\omega_{v,i} \sqrt{\mathcal{B}_{m,i}}}.$$

The number of order terms can be reduced by noting that the leading ones are given by  $O_p(mJT^{-1/2})$  and  $O_p(\sqrt{m}T^{-1/2+1/\kappa})$ . For these to be  $o(1)$ , we require  $m = o(T^{1/2-1/\kappa})$  and  $J = o(mT^{-1/2})$ . Hence, in this case,  $\theta_i = {}_d \mathcal{D}_{m,i} + o_p(1)$ , which, together with the continuous mapping theorem, in turn implies

$$\begin{aligned} \tau_{SQ}(q_{k+1}, q_k) &= {}_d \mathcal{D}_{(p_{k+1}),m} + o_p(1), \\ \tau_{GM} &= {}_d \sum_{i=1}^N \mathcal{D}_{m,i} + o_p(1). \end{aligned} \tag{A.23}$$

As for  $\tau_p$ , by using a similar expansion as the one above for  $\theta_i$ ,

$$\tau_p = {}_d \frac{\sum_{i=1}^N (mb_i \mathcal{B}_{m,i} + \mathcal{A}_{m,i})}{\sqrt{\sum_{i=1}^N \omega_{v,i}^2 \mathcal{B}_{m,i}}} + o_p(1) = \frac{\text{tr}(mb \odot \mathcal{B}_m + \mathcal{A}_m)}{\sqrt{\text{tr}(\Omega_{vv} \odot \mathcal{B}_m)}} + o_p(1), \tag{A.24}$$

where  $\odot$  is the Hadamard product.

In order to obtain the large- $m$  results, we make use of Lemma A.3, from which it follows that if  $b_i = 0$ , then

$$\theta_i =_d \frac{\mathcal{A}_{m,i}}{\omega_{v,i}\sqrt{\mathcal{B}_{m,i}}} + o_p(1) \rightarrow_d \frac{\sqrt{-2c_i}}{\omega_{v,i}\omega_{w,i}} Y_i = X_i \quad (\text{A.25})$$

as  $T, m, J \rightarrow \infty$ , whereas if  $b_i \neq 0$ , then

$$\frac{1}{\sqrt{m}}\theta_i =_d \frac{b_i\sqrt{m\mathcal{B}_{m,i}}}{\omega_{v,i}} + o_p(1) \rightarrow_p \frac{b_i\omega_{w,i}}{\sqrt{-2c_i}\omega_{v,i}}. \quad (\text{A.26})$$

suggesting that  $\theta_i = O_p(\sqrt{m})$ .

The results for  $\tau_{GM}$ ,  $\tau_P$  and  $\tau_{SQ}(q_{k+1}, q_k)$  are easily deduced from that of  $\theta_i$ . We begin with  $\tau_{SQ}(q_{k+1}, q_k)$ . If  $q \geq q_{k+1}$ , then

$$\tau_{SQ}(q_{k+1}, q_k) \rightarrow_d X_{(p_{k+1})} \quad (\text{A.27})$$

as  $m, J, T \rightarrow \infty$ , whereas if  $q < q_{k+1}$ , then  $\tau_{SQ}(q_{k+1}, q_k) = O_p(\sqrt{m})$ . Similarly, if  $q = 0$ , then

$$\tau_{GM} \rightarrow_d \sum_{i=1}^N X_i, \quad (\text{A.28})$$

whereas if  $q > 0$ , then  $\tau_{GM} = O_p(\sqrt{m})$ . As for  $\tau_P$ , we can show that if  $q = 0$ ,

$$\tau_P =_d \frac{\text{tr}(\sqrt{m}\mathcal{A}_m)}{\sqrt{\text{tr}(\Omega_{vv} \odot m\mathcal{B}_m)}} + o_p(1) \rightarrow_d \frac{\sum_{i=1}^N Y_i}{\sqrt{\text{tr}(\Omega_{vv} \odot C \odot \Omega_{ww})}}. \quad (\text{A.29})$$

On the other hand, if  $q > 0$ , then we again have that  $\tau_P = O_p(\sqrt{m})$ . ■

## B Appendix: Proofs for the bootstrap statistics

**Lemma B.1.** *Under Assumptions 1 and 2 we have that*

$$\begin{aligned} J_{cm,T}^*(r) &= T^{-1/2} \chi_{[Tr]}^* =_{d^*} J_{cm}(r) + O_p^*(\ell^{1/2-1/\kappa} T^{-1/2+1/\kappa}) \\ &\quad + O_p^*(\ell^{-1/2}) + O_p^*(mT^{-1}) \text{ in probability.} \end{aligned}$$

**Proof of Lemma B.1.**

Let  $B_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t^*$ . We then have that

$$\begin{aligned} J_{0,T}^*(r) &= \frac{1}{\sqrt{T}} \sum_{m=1}^{M(r)-1} \sum_{s=1}^{\ell} \left( \tilde{u}_{\mathcal{I}_m+s} - \frac{1}{T-\ell} \sum_{\tau=1}^{T-\ell} \tilde{u}_{\tau+s} \right) \\ &\quad + \frac{1}{\sqrt{T}} \sum_{s=1}^{N(r)} \left( \tilde{u}_{\mathcal{I}_{M_r}+s} - \frac{1}{T-\ell} \sum_{\tau=1}^{T-\ell} \tilde{u}_{\tau+s} \right) \\ &= \frac{1}{\sqrt{T}} \sum_{m=1}^{M(r)} \sum_{s=1}^{\ell} \left( \tilde{u}_{\mathcal{I}_m+s} - \frac{1}{T-\ell} \sum_{\tau=1}^{T-\ell} \tilde{u}_{\tau+s} \right) - R_T^*(r), \end{aligned} \tag{B.1}$$

where  $M(r) = \lceil \lfloor Tr \rfloor / \ell \rceil$ ,  $N(r) = \lfloor Tr \rfloor - M(r)\ell - 1$  and  $R_T(r) = T^{-1/2} \sum_{s=N(r)+1}^{\ell} (\tilde{u}_{\mathcal{I}_{M(r)}+s} - (T-\ell)^{-1} \sum_{\tau=1}^{T-\ell} \tilde{u}_{\tau+s})$ . Define  $\gamma = (\beta', \rho')'$  and correspondingly  $\tilde{\gamma} = (\tilde{\beta}', \tilde{\rho}')'$ . Letting  $z_t = (y'_t, x'_t)$ , we have that

$$\begin{aligned} \tilde{u}_{(\mathcal{I}_m+s)} &= z_{\mathcal{I}_m+s}^d - \tilde{\gamma} x_{\mathcal{I}_m+s-1}^d - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} (z_{t+s}^d - \tilde{\gamma} x_{t+s-1}^d) \\ &= z_{\mathcal{I}_m+s} - \tilde{\gamma} x_{\mathcal{I}_m+s-1} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} (z_{t+s} - \tilde{\gamma} x_{t+s-1}) \\ &= u_{\mathcal{I}_m+s} - (\tilde{\gamma} - \gamma) x_{\mathcal{I}_m+s-1} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} (u_{t+s} - (\tilde{\gamma} - \gamma) x_{t+s-1}), \end{aligned}$$

which in turn implies

$$\begin{aligned} R_T^*(r) &= \frac{1}{\sqrt{T}} \sum_{s=N(r)+1}^{\ell} \left( u_{\mathcal{I}_{M(r)}+s} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} u_{t+s} \right) \\ &\quad + T^{-1/2} (\tilde{\gamma} - \gamma) \sum_{s=N(r)+1}^{\ell} \left( x_{\mathcal{I}_{M(r)}+s-1} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} x_{t+s-1} \right) \\ &= R_{1,T}^*(r) - R_{2,T}^*(r), \end{aligned}$$

where  $R_{1,T}^*(r)$  and  $R_{2,T}^*(r)$  are implicitly defined. By using the stationarity of  $u_t$ , we can show that  $R_{1,T}^*(r) = O_p^*(\sqrt{\ell} T^{-1/2})$ . Also,  $R_{2,T}^*(r) = O_p^*(\ell T^{-1})$  as  $\tilde{\gamma} - \gamma = O_p(\sqrt{m} T^{-1})$  and

$x_t = O_p(\sqrt{mT})$ . Hence,  $R_T^*(r) = O_p(\sqrt{\ell}T^{-1/2})$ , which in turn implies

$$\begin{aligned} B_T^*(r) &= \frac{1}{\sqrt{T}} \sum_{m=1}^{M(r)} \sum_{s=1}^{\ell} \left( u_{\mathcal{I}_m+s} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} u_{t+s} \right) \\ &\quad - T^{-1/2}(\tilde{\gamma} - \gamma) \sum_{m=1}^{M(r)} \sum_{s=1}^{\ell} \left( x_{\mathcal{I}_m+s-1} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} x_{t+s-1} \right) + O_p^*(\sqrt{\ell}T^{-1/2}) \quad (\text{B.2}) \\ &= B_{1,T}^*(r) - B_{2,T}^*(r) + O_p^*(\sqrt{\ell}T^{-1/2}), \end{aligned}$$

where  $B_{1,T}^*(r)$  and  $B_{2,T}^*(r)$  are again implicitly defined. Consider  $B_{2,T}^*(r)$ . Let us define  $b_{T,m}^* = \sum_{s=1}^{\ell} (x_{\mathcal{I}_m+s-1} - \mathbb{E}^* x_{\mathcal{I}_m+s-1}) = \sum_{s=1}^{\ell} (x_{\mathcal{I}_m+s-1} - (T-\ell)^{-1} \sum_{\tau=1}^{T-\ell} x_{\tau+s-1})$ . Clearly,  $\mathbb{E}^* b_{T,m}^* = 0$ . Moreover,

$$\begin{aligned} \mathbb{E}^* \|b_{T,m}^*\|^2 &= \mathbb{E}^* \left[ \left( \sum_{s=1}^{\ell} x_{\mathcal{I}_m+s-1} \right)' \left( \sum_{s=1}^{\ell} x_{\mathcal{I}_m+s-1} \right) \right] - \left( \mathbb{E}^* \sum_{s=1}^{\ell} x_{\mathcal{I}_m+s-1} \right)' \left( \mathbb{E}^* \sum_{s=1}^{\ell} x_{\mathcal{I}_m+s-1} \right) \\ &= \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} \left\| \sum_{s=1}^{\ell} x_{t+s-1} \right\|^2 - \left\| \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} \sum_{s=1}^{\ell} x_{t+s-1} \right\|^2 \\ &\leq \frac{\ell}{T-\ell} \sum_{t=1}^{T-\ell} \sum_{s=1}^{\ell} \|x_{t+s-1}\|^2 + O_p(\ell^2 m^{-1} T) = O_p(\ell^2 m^{-1} T). \end{aligned}$$

By using this and the independence of the blocks,

$$\begin{aligned} \mathbb{E}^* \left\| \frac{1}{\sqrt{T}} \sum_{m=1}^{M(r)} b_{T,m}^* \right\|^2 &= \frac{1}{T} \sum_{m_1=1}^{M(r)} \sum_{m_2=1}^{M(r)} \mathbb{E}^* b_{T,m_1}^{*'} b_{T,m_2}^* \\ &= \frac{1}{T} \sum_{m=1}^{M(r)} \mathbb{E}^* \|b_{T,m}^*\|^2 = O_p(k\ell^2 m^{-1}) = O_p(\ell m^{-1} T). \end{aligned}$$

It follows that

$$B_{2,T}^*(r) = O_p^*(\ell^{1/2} T^{-1/2}). \quad (\text{B.3})$$

Next, consider  $B_{1,T}^*(r)$ . Note that  $B_{1,T}^*(r) = T^{-1/2} \sum_{m=1}^{M(r)} \sum_{s=1}^{\ell} (u_{\mathcal{I}_m+s} - \mathbb{E}^* u_{\mathcal{I}_m+s})$ . As  $M(r)$  is asymptotically equivalent to  $\lfloor kr \rfloor$ , we can conclude that

$$B_T^*(r) = \frac{1}{\sqrt{T}} \sum_{m=1}^{\lfloor kr \rfloor} \sum_{s=1}^{\ell} (u_{\mathcal{I}_m+s} - \mathbb{E}^* u_{\mathcal{I}_m+s}) + O_p^*(\sqrt{\ell}T^{-1/2}). \quad (\text{B.4})$$

Now let  $E_m^* = \ell^{-1/2} \sum_{s=1}^{\ell} (\varepsilon_{\mathcal{I}_m+s} - \mathbb{E}^* \varepsilon_{\mathcal{I}_m+s})$ . A modification of Lemma A.3 of Palm et al. (2011) gives

$$\Sigma^* = \mathbb{E}^* E_m^* E_m^{*'} = \frac{1}{\ell(T-\ell)} \sum_{s_1=1}^{\ell} \sum_{s_2=1}^{\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s_1} \varepsilon'_{t+s_2} + O_p(\ell T^{-1}),$$

where

$$\begin{aligned}
\frac{1}{\ell(T-\ell)} \sum_{s_1=1}^{\ell} \sum_{s_2=1}^{\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s_1} \varepsilon'_{t+s_2} &= \frac{1}{\ell(T-\ell)} \sum_{s=1}^{\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s} \varepsilon'_{t+s} \\
&\quad + \frac{1}{\ell(T-\ell)} \sum_{s_1=1}^{\ell} \sum_{s_2=1, s_1 \neq s_2}^{\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s_1} \varepsilon'_{t+s_2} \\
&= \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon'_t + O_p((\ell T)^{-1/2}) = \Sigma_{\varepsilon\varepsilon} + O_p((\ell T)^{-1/2}),
\end{aligned}$$

which holds as  $\mathbb{E} \left\| \sum_{s=1}^{\ell} \sum_{j=s+1}^{\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s} \varepsilon'_{t+j} \right\|^2 = O(\ell^3 T)$ . Hence,

$$\Sigma^* = \mathbb{E}^* E_m^* E_m^{*'} = \Sigma + O_p(\sqrt{\ell} T^{-1/2}) \quad (\text{B.5})$$

Let us denote by  $B_{\varepsilon}(r)$  a Brownian motion with covariance matrix  $\Sigma_{\varepsilon\varepsilon}$  and let  $B_{\varepsilon,k}^*(r) = k^{-1/2} \sum_{m=1}^{\lfloor kr \rfloor} E_m^*$ . Then, following Chang et al. (2006, page 714) (also see Park, 2006, Proof of Lemma 2.4), we have that

$$\mathbb{P}^* \left\{ \sup_{0 \leq r \leq 1} |B_{\varepsilon,k}^*(r) - B_{\varepsilon}(r)| > k^{-1/2} c_k \right\} \leq K k c_k^{-\kappa} \mathbb{E}^* |E_m^*|^{\kappa}.$$

Now take  $c_k = k^{1/\kappa} M$  for a large  $M > 0$  to obtain

$$\mathbb{P}^* \left\{ \sup_{0 \leq r \leq 1} |B_{\varepsilon,k}^*(r) - B_{\varepsilon}(r)| > k^{-1/2+1/\kappa} \right\} \leq K M^{-\kappa} \mathbb{E}^* |E_m^*|^{\kappa}. \quad (\text{B.6})$$

We now show that  $\mathbb{E}^* |E_m^*|^{\kappa} = O_p(1)$ . By the definition of  $E_m^*$ ,

$$\begin{aligned}
\mathbb{E}^* |E_m^*|^{\kappa} &= \frac{1}{T-\ell} \sum_{t=1}^T \left| \frac{1}{\sqrt{\ell}} \sum_{s=1}^{\ell} \left( \varepsilon_{t+s} - \frac{1}{T-\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s} \right) \right|^{\kappa} \\
&\leq 2^{\kappa-1} \frac{1}{T-\ell} \sum_{t=1}^T \left[ \left| \frac{1}{\sqrt{\ell}} \sum_{s=1}^{\ell} \varepsilon_{t+s} \right|^{\kappa} + \left| \frac{1}{\sqrt{\ell}(T-\ell)} \sum_{s=1}^{\ell} \sum_{t=1}^{T-\ell} \varepsilon_{t+s} \right|^{\kappa} \right] \\
&= 2^{\kappa-1} (E_{1,T} + E_{2,T}),
\end{aligned}$$

with implicit definitions of  $E_{1,T}$  and  $E_{2,T}$ .  $E_{2,T}$  is clearly  $o_p(1)$ . As for  $E_{1,T}$ , note that for any  $\delta > 0$ ,

$$\mathbb{P}(E_{1,T} > \delta) \leq \delta^{-1} \mathbb{E} E_{1,T} = \delta^{-1} \mathbb{E} \left| \frac{1}{\sqrt{\ell}} \sum_{s=1}^{\ell} \varepsilon_{t+s} \right|^{\kappa}.$$

Successive application of the Marcinkiewicz-Zygmund and Minkowski inequalities element-

by-element yields

$$\begin{aligned}
\mathbb{E} \left| \frac{1}{\sqrt{\ell}} \sum_{s=1}^{\ell} \varepsilon_{t+s} \right|^\kappa &= \ell^{-\kappa/2} \mathbb{E} \left[ \sum_{i=1}^N \left( \sum_{s=1}^{\ell} \varepsilon_{i,t+s} \right)^2 \right]^{\kappa/2} \\
&\leq \ell^{-\kappa/2} N^{\kappa/2-1} \sum_{i=1}^N \mathbb{E} \left| \sum_{s=1}^{\ell} \varepsilon_{i,t+s} \right|^\kappa \leq c_\kappa \ell^{-\kappa/2} N^{\kappa/2-1} \sum_{i=1}^N \mathbb{E} \left( \sum_{s=1}^{\ell} \varepsilon_{i,t+s}^2 \right)^{\kappa/2} \\
&\leq c_\kappa \ell^{-\kappa/2} N^{\kappa/2-1} \sum_{i=1}^N \left[ \sum_{s=1}^{\ell} \left\{ \mathbb{E} |\varepsilon_{i,t+s}|^\kappa \right\}^{2/\kappa} \right]^{\kappa/2} \\
&= c_\kappa \ell^{-\kappa/2} N^{\kappa/2-1} \sum_{i=1}^N \left[ \ell \left\{ \mathbb{E} |\varepsilon_{i,t}|^\kappa \right\}^{2/\kappa} \right]^{\kappa/2} \\
&= c_\kappa N^{\kappa/2-1} \sum_{i=1}^N \mathbb{E} |\varepsilon_{i,t}|^\kappa \leq C_\kappa \mathbb{E} |\varepsilon_t|^\kappa,
\end{aligned}$$

where  $c_\kappa$  and  $C_\kappa$  do not depend on  $\ell$ . It follows that  $\mathbb{E}^* |E_m^*|^\kappa = O_p(1)$ , and therefore

$$B_{\varepsilon,k}^*(r) =_d B_\varepsilon(r) + O_p^*(k^{-1/2+1/\kappa}) \text{ in probability.} \quad (\text{B.7})$$

By using this and Lemma A.5 of Palm et al. (2011),

$$\frac{1}{\sqrt{T}} \sum_{m=1}^{\lfloor kr \rfloor} \sum_{s=1}^{\ell} (u_{\mathcal{I}_m+s} - \mathbb{E}^* u_{\mathcal{I}_m+s}) = \Psi(1) B_{\varepsilon,k}^*(r) + O_p^*(\ell^{-1/2}). \quad (\text{B.8})$$

Putting all results together, we get that

$$B_T^*(r) =_d B(r) + O_p^*(\ell^{-1/2}) + O_p^* \left( (\ell T^{-1})^{1/2-1/\kappa} \right) \text{ in probability,} \quad (\text{B.9})$$

where  $B(r)$  is as in Lemma A.1. As  $\tilde{\rho}_i = \rho_i + O_p(\sqrt{m}T^{-1}) = 1 + T^{-1}c_i m + o_p(mT^{-1})$ , the result follows directly from the proof of Lemma 2.3 and 2.4 in Park (2006). ■

**Lemma B.2.** *Under Assumptions 1–4 and B.1, we have, uniformly in  $m$ ,*

$$\begin{aligned}
\frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^{*d} v_t^{*'} &=_{d^*} \sqrt{m} (\mathcal{A}_m - \Lambda_{wv}^*) + O_p^*(R_{mT}^*) \text{ in probability,} \\
\frac{m}{T^2} \sum_{t=2}^T x_{t-1}^{*d} x_{t-1}^{*d'} &=_{d^*} m \mathcal{B}_m + O_p^*(R_{mT}^*) \text{ in probability,}
\end{aligned}$$

where  $R_{mT}^* = \sqrt{m} \ell^{1/2-1/\kappa} T^{-1/2+1/\kappa} + m^{3/2} T^{-1} + \sqrt{m} \ell^{-1/2}$ .

**Proof of Lemma B.2.**

Let  $\mathcal{F}_t^*$  denote the smallest sigma-field containing all past information up to time  $t$  on  $x_t^*$  and

$v_t^*$ , and write  $\mathbb{E}_t^* X = \mathbb{E}(X|\mathcal{F}_t^*)$  for the conditional expectation of  $X$  at time  $t$ . Now we can construct a martingale difference sequence  $\varepsilon_t^*$  with respect to  $\mathcal{F}_t^*$  in the same way as Hansen (1992, Section 3):

$$\varepsilon_t^* = \sum_{j=0}^{\infty} (\mathbb{E}_t^* v_{t+j}^* - \mathbb{E}_{t-1}^* v_{t+j}^*),$$

Let  $t = (k_t - 1)\ell + s_t$ . By the block-wise nature of the bootstrap sample and the independence of the blocks, the conditional expectation of a variable in a different (future) block is equal to the unconditional bootstrap expectation, while that of a variable within the block is the variable itself. Hence,

$$\mathbb{E}_{(k_t-1)\ell+s_t}^* v_{t+j}^* = \begin{cases} \mathbb{E}^* v_{(k_t-1)\ell+s_t+j}^* = \mathbb{E}^* v_{t+k}^* = 0 & \text{if } j > l - s_t \\ \mathbb{E}^* v_{(k_t-1)\ell+s_t+j}^* = v_{(k_t-1)\ell+s_t+j}^* = \hat{v}_{\mathcal{I}_{k_t-1}+s_t+j} & \text{if } j \leq l - s_t \end{cases}.$$

Using that  $\mathbb{E}_t^* v_{t+j}^* = \mathbb{E}_{t-1}^* v_{t+j}^*$  for any  $j$  if  $t$  and  $t-1$  are in the same block, we can further show that

$$\begin{aligned} \varepsilon_t^* &= \varepsilon_{(k_t-1)\ell+s_t}^* = \sum_{j=0}^{\ell-s_t} (\mathbb{E}_{(k_t-1)\ell+s_t}^* v_{(k_t-1)\ell+s_t+j}^* - \mathbb{E}_{(k_t-1)\ell+s_t}^* v_{(k_t-1)\ell+s_t+j}^*) \\ &= \begin{cases} \sum_{j=1}^{\ell} (v_{(k_t-1)\ell+j}^* - \mathbb{E}^* v_{(k_t-1)\ell+j}^*) = \sum_{j=1}^{\ell} v_{(k_t-1)\ell+j}^* & \text{if } s_t = 1 \\ 0 & \text{if } s_t > 1 \end{cases}. \end{aligned}$$

Define  $Y_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t^*$ . It follows directly from the proof of Lemma B.1 that

$$Y_T^*(r) = \frac{1}{\sqrt{T}} \sum_{m=1}^{M(r)} \sum_{j=1}^{\ell} \hat{v}_{(\mathcal{I}_m-1)\ell+j} \rightarrow_{d^*} B_v(r) \text{ in probability,} \quad (\text{B.10})$$

where  $B_v(r)$  is as in Lemma A.1. Application of Theorem 2.1 of Hansen (1992) now yields, as  $T \rightarrow \infty$ ,

$$\frac{1}{T} \sum_{t=2}^T x_{t-1}^{*d} \varepsilon_t^* = \int_0^1 J_{cm,T}^{*d}(r) dY_T^*(r) \rightarrow_{d^*} \int_0^1 J_{cm}^d(r) dB_v(r) \text{ in probability.} \quad (\text{B.11})$$

To obtain the appropriate approximation orders, we proceed as in the proof of Lemma 3.1 in Park (2006), and embed the partial sum of  $\varepsilon_t^*$  into a Brownian motion with properly chosen stopping times. This gives

$$\sqrt{m} \int_0^1 J_{cm,T}^{*d}(r) dY_T^*(r) =_{d^*} \sqrt{m} \int_0^1 J_{cm,T}^{*d}(r) dB_v(r) + O_p^*(\sqrt{m}T^{-1/2}) \text{ in probability.} \quad (\text{B.12})$$

Clearly,

$$\begin{aligned} \sqrt{m} \int_0^1 J_{cm,T}^{*d}(r) dB_v(r) &=_{d^*} \sqrt{m} \int_0^1 J_{cm}^d(r) dB_v(r) \\ &\quad + \sqrt{m} \int_0^1 (J_{cm,T}^{*d}(r) - J_{cm}^d(r)) dB_v(r) \text{ in probability,} \end{aligned}$$

which, by using the same steps as in the proof of Lemma 3.1 in Park (2006), gives

$$\begin{aligned} \sqrt{m} \int_0^1 J_{cm,T}^{*d}(r) dY_T^*(r) &=_{d^*} \sqrt{m} \mathcal{A}_m + O_p^*(\sqrt{m} \ell^{1/2-1/\kappa} T^{-1/2+1/\kappa}) + O_p^*(\sqrt{m} \ell^{-1/2}) \\ &\quad + O_p^*(m^{3/2} T^{-1}) \text{ in probability,} \end{aligned}$$

where the orders follow from Lemma B.1.

Now define  $\zeta_t^*$  recursively as  $\Delta \zeta_t^* = \varepsilon_t^* - v_t^*$ , where  $\zeta_0^* = 0$ . Clearly,

$$\begin{aligned} \zeta_t^* &= \zeta_{(k_t-1)\ell+s_t}^* = \sum_{m=1}^{k_t-1} \sum_{j=1}^{\ell} \Delta \zeta_{(m-1)\ell+j}^* + \sum_{j=1}^{s_t} \Delta \zeta_{(k_t-1)\ell+j}^* \\ &= \sum_{m=1}^{k_t-1} \left( \sum_{j=1}^{\ell} v_{(m-1)\ell+j}^* - \sum_{j=1}^{\ell} v_{(m-1)\ell+j}^* \right) + \sum_{j=1}^{\ell} v_{(k_t-1)\ell+j}^* - \sum_{j=1}^{s_t} v_{(k_t-1)\ell+j}^* \\ &= \sum_{j=s_t+1}^{\ell} v_{(k_t-1)\ell+j}^*, \end{aligned}$$

suggesting that

$$\frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^{*d} v_t^* = \frac{\sqrt{m}}{T} \sum_{t=2}^T x_{t-1}^{*d} \varepsilon_t^* - \sqrt{m} \Lambda_T^*, \quad (\text{B.13})$$

where

$$\begin{aligned} \Lambda_T^* &= -\frac{1}{T} \sum_{t=2}^T x_{t-1}^{*d} \Delta \zeta_t^* = \frac{1}{T} \sum_{t=2}^{T-1} \zeta_t^* (x_t^{*d} - x_{t-1}^{*d}) - T^{-1} (\zeta_1^* x_1^{*d} - \zeta_T^* x_T^{*d}) \\ &= \frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} (x_{(m-1)\ell+s}^{*d} - x_{(m-1)\ell+s-1}^{*d}) \left( \sum_{j=s+1}^{\ell} v_{(m-1)\ell+j}^* \right) + O_p^*(\sqrt{\ell}(mT)^{-1/2}). \end{aligned}$$

Substitution of  $x_t^* - x_{t-1}^* = (\tilde{\rho} - I_N) x_{t-1}^* + w_t^*$  yields

$$\begin{aligned} \Lambda_T^* &= T^{-1} (\tilde{\rho} - I_N) \sum_{m=1}^k \sum_{s=1}^{\ell} x_{(m-1)\ell+s-1}^{*d} \left( \sum_{j=s+1}^{\ell} v_{(m-1)\ell+j}^* \right)' \\ &\quad + \frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} w_{(m-1)\ell+s}^* \left( \sum_{j=s+1}^{\ell} v_{(m-1)\ell+j}^* \right)' + o_p^*(1) = \Lambda_{1,T}^* + \Lambda_{2,T}^*, \end{aligned} \quad (\text{B.14})$$

with implicit definitions of  $\Lambda_{1,T}^*$  and  $\Lambda_{2,T}^*$ . Now, since  $(\tilde{\rho} - I_N) = O_p(mT^{-1})$ , we have  $\Lambda_{1,T}^* = O_p^*(\sqrt{m\ell}T^{-1})$ . It remains to show that  $\Lambda_{2,T}^* \rightarrow_{p^*} \Lambda_{wv}$  in probability. For this purpose it is convenient to define  $D_m^* = \ell^{-1} (\sum_{s=1}^{\ell} u_{(m-1)\ell+s}^*) (\sum_{s=1}^{\ell} u_{(m-1)\ell+s}^{*'})$ ,  $D_t = (\sum_{s=1}^{\ell} u_{t+s}^*) (\sum_{s=1}^{\ell} u_{t+s}^{*'})$  and  $\Omega_T^* = k^{-1} \sum_{m=1}^k D_m^*$ . Note first that due to the independence of the blocks,

$$\mathbb{P}^* (\|\Omega_T^* - \mathbb{E}^* \Omega_T^*\| > \epsilon) \leq \epsilon^{-2} k^{-2} \sum_{m=1}^k \mathbb{E}^* \|D_m^* - \mathbb{E}^* D_m^*\|^2,$$



where, by the linear process assumption,  $\mathbb{E}^* \|D_m^*\|^2 = (T - \ell)^{-1} \sum_{t=1}^{T-\ell} \|D_t^2\| = O_p(1)$ , and therefore,

$$\Omega_T^* = \mathbb{E}^* \Omega_T^* + O_p^*(k^{-1/2}),$$

where, by the results provided in the proof of Lemma B.1,

$$\mathbb{E}^* \Omega_T^* = \mathbb{E}^* B_T^*(1) B_T^*(1)' = \Omega + O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{\ell} T^{-1/2}) \text{ in probability.}$$

By similar arguments,

$$T^{-1} \sum_{m=1}^k \sum_{s=1}^{\ell} u_{(m-1)\ell+s}^* u_{(m-1)\ell+j}^{*'} = \Sigma + O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{\ell} T^{-1/2}) \text{ in probability.}$$

The previous two results imply

$$T^{-1} \sum_{m=1}^k \sum_{s=1}^{\ell} u_{(m-1)\ell+s}^* \left( \sum_{j=s+1}^{\ell} u_{(m-1)\ell+j}^* \right)' = \Lambda + O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{\ell} T^{-1/2}) \text{ in probability,}$$

and therefore

$$B_{2,T}^* = \Lambda_{wv} + O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{\ell} T^{-1/2}) \text{ in probability.} \quad (\text{B.15})$$

Hence,

$$\Lambda_T^* = \Lambda_{wv} + O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{\ell} T^{-1/2}) + O_p^*(\sqrt{m\ell} T^{-1}). \quad (\text{B.16})$$

and so the proof is complete. ■

**Lemma B.3.** *Under the conditions of Lemma B.1*

$$\|\hat{\Lambda}^* - \Lambda\| = O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{m\ell} T^{-1/2}) \text{ in probability,}$$

$$\|\hat{\Omega}^* - \Omega\| = O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{m\ell} T^{-1/2}) \text{ in probability.}$$

**Proof of Lemma B.3.**

Since the contribution of the final (partial) block is of order  $O_p^*(\ell T^{-1})$ , we may assume that

$T = k\ell$ . Note that, since  $\hat{u}_t^* = u_t^* - (\hat{\gamma}^* - \gamma^*)x_{t-1}^{*d}$ ,

$$\begin{aligned}
& \frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} \hat{u}_{(m-1)\ell+s}^* \hat{u}_{(m-1)\ell+j}^{*'} \\
&= \frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} u_{(m-1)\ell+s}^* u_{(m-1)\ell+j}^{*'} \\
&\quad + T^{-1} (\hat{\gamma}^* - \gamma^*) \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} x_{(m-1)\ell+s-1}^{*d} u_{(m-1)\ell+j}^{*'} \\
&\quad + T^{-1} \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} u_{(m-1)\ell+s}^* x_{(m-1)\ell+j-1}^{*d'} (\hat{\gamma}^* - \gamma^*)' \\
&\quad + T^{-1} (\hat{\gamma}^* - \gamma^*) \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} x_{(m-1)\ell+s-1}^{*d} x_{(m-1)\ell+j-1}^{*d'} (\hat{\gamma}^* - \gamma^*)'.
\end{aligned}$$

By the Cauchy-Schwartz inequality the second term on the right-hand side can be written as

$$\begin{aligned}
& \left\| T^{-1} (\hat{\gamma}^* - \gamma^*) \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} x_{(m-1)\ell+s-1}^{*d} u_{(m-1)\ell+j}^{*'} \right\| \\
& \leq T^{-1} \|\hat{\gamma}^* - \gamma^*\| \sum_{s=1}^{\ell} \left( \sum_{j=1}^{\ell} \sum_{m=1}^k \left\| x_{(m-1)\ell+s-1}^{*d} \right\|^2 \right)^{1/2} \left( \sum_{m=1}^k \left\| u_{(m-1)\ell+j}^{*'} \right\|^2 \right)^{1/2},
\end{aligned}$$

which is  $O_p^*(\sqrt{m\ell}T^{-1/2})$ , as  $x_t^* = O_p^*(m^{-1/2}\sqrt{T})$  and  $\|\hat{\gamma} - \gamma\| = O_p^*(mT^{-1})$  (see Lemmas B.1 and B.2, respectively). The same reasoning can be applied to show that the third and fourth terms are  $O_p^*(\sqrt{m\ell}T^{-1/2})$  and  $O_p^*(m\ell T^{-1})$ , respectively. Therefore,

$$\frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} \hat{u}_{(m-1)\ell+s}^* \hat{u}_{(m-1)\ell+j}^{*'} = \frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} u_{(m-1)\ell+s}^* u_{(m-1)\ell+j}^{*'} + O_p^*(\sqrt{m\ell}T^{-1/2}).$$

It follows directly from the proof of Lemma B.2 that

$$\frac{1}{T} \sum_{m=1}^k \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} u_{(m-1)\ell+s}^* u_{(m-1)\ell+j}^{*'} = \Omega + O_p^*(\ell^{-1/2}) + O_p^*(\sqrt{\ell}T^{-1/2}) \text{ in probability. (B.17)}$$

The proof for  $\|\hat{\Lambda}^* - \Lambda\|$  is analogous. ■

### Proof of Theorem 2.

The proof follows from Lemmas B.1 and B.2, and by using the same steps as in the proof of Lemma A.2 and Theorem 1. ■

### Proof of Corollary 1.

The proof follows directly from that of Corollary 1 in Smeekees (2011), and is therefore omitted. ■

Table 1: Empirical rejection frequencies of  $\tau_P$  and  $\tau_{GM}$ .

$\gamma$	$N$	$T$	Case	$q = 0$		$q = 0.2$		$q = 0.5$		$q = 0.9$	
				$\tau_P$	$\tau_{GM}$	$\tau_P$	$\tau_{GM}$	$\tau_P$	$\tau_{GM}$	$\tau_P$	$\tau_{GM}$
0.9	10	100	1	0.051	0.040	0.175	0.113	0.836	0.808	0.999	1.000
			2	0.053	0.038	0.157	0.077	0.889	0.885	0.998	0.998
			3	0.069	0.059	0.487	0.484	0.991	0.982	1.000	1.000
		250	1	0.061	0.042	0.370	0.341	0.838	0.894	1.000	1.000
			2	0.048	0.035	0.544	0.504	0.956	0.951	1.000	1.000
			3	0.068	0.068	0.837	0.774	1.000	0.997	1.000	1.000
	25	100	1	0.035	0.027	0.663	0.494	1.000	1.000	1.000	1.000
			2	0.039	0.021	0.611	0.480	0.999	0.994	1.000	1.000
			3	0.052	0.048	0.825	0.714	1.000	1.000	1.000	1.000
		250	1	0.055	0.044	0.711	0.664	1.000	1.000	1.000	1.000
			2	0.052	0.048	0.744	0.635	1.000	1.000	1.000	1.000
			3	0.073	0.055	0.884	0.854	1.000	1.000	1.000	1.000
1	10	100	1	0.049	0.040	0.062	0.063	0.706	0.688	0.989	0.977
			2	0.044	0.031	0.201	0.148	0.588	0.535	0.993	0.981
			3	0.082	0.045	0.716	0.538	0.890	0.855	0.998	0.997
		250	1	0.059	0.049	0.133	0.105	0.790	0.755	0.989	0.987
			2	0.039	0.035	0.188	0.148	0.749	0.660	0.984	0.976
			3	0.073	0.063	0.676	0.593	0.956	0.942	1.000	1.000
	25	100	1	0.039	0.031	0.349	0.237	0.980	0.959	1.000	1.000
			2	0.033	0.025	0.450	0.348	0.996	0.981	1.000	1.000
			3	0.067	0.044	0.858	0.797	0.996	0.993	1.000	1.000
		250	1	0.036	0.040	0.438	0.361	0.985	0.976	1.000	1.000
			2	0.038	0.047	0.311	0.304	0.971	0.971	1.000	1.000
			3	0.083	0.051	0.797	0.801	0.991	0.993	1.000	1.000

Notes:  $q$  refers to the fraction of predictable units and  $\gamma$  is such that  $m = T^{1-\gamma}$ . Cases 1–3 refer to the extent of serial and cross-sectional correlation, and endogeneity.

Table 2: Average proportions of correctly and incorrectly selected predictable units using  $\tau_{SEQ}$ .

$\gamma$	$N$	$T$	Case	$q = 0$		$q = 0.2$		$q = 0.5$		$q = 0.9$	
				ICP	CP	ICP	CP	ICP	CP	ICP	CP
0.9	10	100	1	0.001	–	0.001	0.091	0.001	0.149	0.004	0.186
			2	0.001	–	0.001	0.037	0.001	0.210	0.004	0.165
			3	0.002	–	0.001	0.487	0.004	0.596	0.019	0.523
		250	1	0.003	–	0.004	0.501	0.005	0.312	0.004	0.508
			2	0.004	–	0.003	0.675	0.003	0.369	0.008	0.501
			3	0.007	–	0.006	0.589	0.004	0.655	0.085	0.864
	25	100	1	0.000	–	0.000	0.006	0.000	0.001	0.000	0.003
			2	0.000	–	0.000	0.001	0.000	0.000	0.000	0.002
			3	0.000	–	0.000	0.090	0.000	0.192	0.000	0.076
		250	1	0.000	–	0.000	0.122	0.000	0.130	0.000	0.098
			2	0.000	–	0.000	0.076	0.000	0.127	0.000	0.119
			3	0.000	–	0.000	0.261	0.000	0.264	0.001	0.490
1	10	100	1	0.001	–	0.001	0.025	0.002	0.114	0.000	0.092
			2	0.002	–	0.001	0.159	0.001	0.066	0.000	0.106
			3	0.001	–	0.002	0.546	0.002	0.304	0.009	0.346
		250	1	0.005	–	0.005	0.069	0.004	0.214	0.004	0.174
			2	0.003	–	0.004	0.137	0.003	0.156	0.004	0.138
			3	0.012	–	0.007	0.496	0.011	0.717	0.047	0.576
	25	100	1	0.000	–	0.000	0.000	0.000	0.000	0.000	0.000
			2	0.000	–	0.000	0.000	0.000	0.000	0.000	0.000
			3	0.000	–	0.000	0.236	0.000	0.074	0.000	0.061
		250	1	0.000	–	0.000	0.021	0.000	0.034	0.000	0.025
			2	0.000	–	0.000	0.017	0.000	0.018	0.000	0.032
			3	0.001	–	0.002	0.490	0.000	0.276	0.000	0.358

Notes: CP and ICP refer to the average proportion of units correctly and incorrectly classified as predictable, respectively. See Table 1 for an explanation of the rest.